

---

# The Shrinkage-Delinkage Trade-off: An Analysis of Factorized Gaussian Approximations for Variational Inference (Supplementary Material)

---

Charles C. Margossian<sup>1</sup>

Lawrence K. Saul<sup>1</sup>

<sup>1</sup>Center for Computational Mathematics, Flatiron Institute, New York, NY, USA

In appendix A, we provide further details on FG-VI when the covariance matrix has constant off-diagonal terms. Next appendix B provides an algorithm to efficiently compute bounds on the shrinkage and delinkage terms, using techniques developed in section 4. In appendix C, we show how to derive bounds on the trace of the shrinkage matrix. In appendix D, we obtain an upper bound on the KL divergence—equivalently the entropy gap—between  $q$  and  $p$ .

## A FG-VI FOR COVARIANCE WITH CONSTANT OFF-DIAGONAL TERMS

Let  $p$  be a multivariate Gaussian distribution over  $\mathbb{R}^n$  with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . The elements of the correlation matrix are related to those of the covariance matrix by

$$C_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}. \quad (1)$$

In this section we assume that the correlation matrix has constant off-diagonal terms, and we use  $\varepsilon \in [0, 1)$  to denote the value of these terms. Note that we require  $\varepsilon \geq 0$  since three or more random variables cannot all be mutually anti-correlated. Also we require  $\varepsilon < 1$  since otherwise  $\mathbf{C}$  (and hence  $\boldsymbol{\Sigma}$ ) would not be positive-definite.

### A.1 SOLUTION FOR FACTORIZED GAUSSIAN VARIATIONAL INFERENCE

Let  $q$  be the solution of FG-VI with diagonal covariance matrix  $\boldsymbol{\Psi}$ , where

$$\Psi_{ii} = \frac{1}{\Sigma_{ii}^{-1}} \quad (2)$$

as in eq. (4).

We now prove Theorem 3.5, broken up into a statement about the estimated variance (Proposition A.1) and a statement about the entropy gap (Proposition A.2).

**Proposition A.1.** *If the correlation matrix in eq. (1) has constant off-diagonal terms, then the solution for FG-VI in eq. (2) obeys the following limits:*

$$\lim_{n \rightarrow \infty} \Psi_{ii} = (1 - \varepsilon)\Sigma_{ii} \quad (3)$$

$$\lim_{\varepsilon \rightarrow 1} \Psi_{ii} = 0 \quad (4)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{trace}(\mathbf{S}) = \frac{1}{1 - \varepsilon} \quad (5)$$

where  $\varepsilon \in [0, 1)$  denotes the value of  $C_{ij}$  for  $i \neq j$ .

*Proof.* Let  $\mathbf{1} \in \mathbb{R}^n$  denote the vector of all ones. Then the correlation matrix can be written as

$$\mathbf{C} = (1 - \varepsilon)\mathbf{I} + \varepsilon\mathbf{1}\mathbf{1}^\top. \quad (6)$$

One can verify by direct substitution that the inverse correlation matrix has elements

$$\mathbf{C}^{-1} = \frac{1}{1 - \varepsilon} \left[ \mathbf{I} - \frac{\varepsilon}{1 + (n-1)\varepsilon} \mathbf{1}\mathbf{1}^\top \right]. \quad (7)$$

Recall from eq. (15) that  $\Psi_{ii} = \Sigma_{ii}/C_{ii}^{-1}$ . With some algebra, it follows from eq. (7) that

$$\Psi_{ii} = \left[ \frac{(1 - \varepsilon)(1 + (n-1)\varepsilon)}{1 + (n-2)\varepsilon} \right] \Sigma_{ii}. \quad (8)$$

It is straightforward to take the limits of eq. (8) as  $n \rightarrow \infty$  or  $\varepsilon \rightarrow 1$ , and these limits yield the results of the proposition.

Finally, recalling that  $\text{trace}(\mathbf{S}) = \sum_{i=1}^n \Sigma_{ii}/\Psi_{ii}$ , we immediately get

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{trace}(\mathbf{S}) = \frac{1}{1 - \varepsilon}. \quad (9)$$

□

**Proposition A.2.** *If the correlation matrix in eq. (1) has constant off-diagonal terms, then the per-component entropy gap from FG-VI vanishes in the limit  $n \rightarrow \infty$ ; that is,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} [\mathcal{H}(p) - \mathcal{H}(q)] = 0. \quad (10)$$

*Proof.* Recall from Theorem 3.2 of the main paper that the entropy gap for FG-VI is given by

$$\mathcal{H}(p) - \mathcal{H}(q) = \frac{1}{2} \log |\mathbf{S}| + \frac{1}{2} \log |\mathbf{C}|, \quad (11)$$

where  $\mathbf{S}$  is the diagonal shrinkage matrix with elements  $S_{ii} = \Sigma_{ii}/\Psi_{ii}$ . We consider each term on the right side of this equation in turn. It follows at once from eq. (8) that

$$\log |\mathbf{S}| = n \left[ \log \frac{1 + (n-2)\varepsilon}{(1-\varepsilon)(1 + (n-1)\varepsilon)} \right], \quad (12)$$

where  $\varepsilon > 0$  denotes the amount of off-diagonal correlation. Next we show how to evaluate  $\log |\mathbf{C}|$ . From eq. (6), we rewrite the correlation matrix as

$$\mathbf{C} = (1-\varepsilon)\mathbf{I} + n\varepsilon \left( \frac{1}{\sqrt{n}}\mathbf{1} \right) \left( \frac{1}{\sqrt{n}}\mathbf{1} \right)^T. \quad (13)$$

Note that the second term on the right side of eq. (13) is a rank-one matrix whose one nonzero eigenvalue is equal to  $n\varepsilon$ . By adding the first term—which is a multiple of the identity matrix—we obtain a new matrix whose eigenvalues are shifted by a uniform amount. It follows that this new matrix (namely,  $\mathbf{C}$ ) has  $n-1$  eigenvalues at  $1-\varepsilon$  and one eigenvalue at  $1 + (n-1)\varepsilon$ , so that

$$\log |\mathbf{C}| = (n-1) \log(1-\varepsilon) + \log(1 + (n-1)\varepsilon). \quad (14)$$

The entropy gap is related to the sum of  $\log |\mathbf{S}|$  and  $\log |\mathbf{C}|$  by eq. (11). Adding the results in eq. (12) and eq. (14), we find that

$$\begin{aligned} \log |\mathbf{S}| + \log |\mathbf{C}| &= -\log(1-\varepsilon) + \log(1 + (n-1)\varepsilon) \\ &\quad + n \log \left[ \frac{1+(n-2)\varepsilon}{1+(n-1)\varepsilon} \right]. \end{aligned} \quad (15)$$

Note that the first term on the right side is  $\mathcal{O}(1)$ , the second term is  $\mathcal{O}(\log n)$ , and the third term can be written as

$$n \log \left[ \frac{1+(n-2)\varepsilon}{1+(n-1)\varepsilon} \right] = n \log \left[ 1 - \frac{\varepsilon}{1+(n-1)\varepsilon} \right]. \quad (16)$$

For large  $n$ , the log term in this equation is  $\mathcal{O}(\frac{1}{n})$  so that the entire expression is  $\mathcal{O}(1)$ . From eq.(15), it therefore follows that the entropy gap in eq. (11) is  $\mathcal{O}(\log n)$ . Dividing by  $n$ , we see that the per-component entropy gap vanishes in the limit  $n \rightarrow \infty$ , thereby completing the proof.  $\square$

From (15), we also see that the entropy gap becomes infinite as  $\varepsilon \rightarrow 1$  (for fixed  $n$ ). Additional limits can be considered with respect to both  $\varepsilon$  and  $n$ , but we do not pursue those here.

## A.2 SOLUTION WHEN MINIMIZING THE REVERSE KL-DIVERGENCE

A factorized approximation cannot both match the marginal variances and the entropy of the target distribution. In the

example at hand, minimizing  $\text{KL}(p||q)$  leads to good estimates of the entropy but not of the marginal variances. We now show that when minimizing the reverse KL-divergence,  $\text{KL}(p||q)$ , the opposite behavior occurs.

We first state the solution obtained when minimizing the reverse KL-divergence. The following is the counterpart to Proposition 2.1 in the main body of the paper.

**Proposition A.3.** *Let  $q(\mathbf{z})$  be multivariate Gaussian with mean  $\tilde{\boldsymbol{\nu}}$  and diagonal covariance  $\tilde{\boldsymbol{\Psi}}$ . Then the variational parameters minimizing the reverse KL-divergence,  $\text{KL}(p||q)$  are given by  $\tilde{\boldsymbol{\nu}} = \boldsymbol{\mu}$  and*

$$\tilde{\Psi}_{ii} = \Sigma_{ii}. \quad (17)$$

*Proof.* The variational parameters  $\tilde{\boldsymbol{\nu}}$  and  $\tilde{\boldsymbol{\Psi}}$  are estimated by minimizing the reverse KL-divergence

$$KL(p||q) = \mathbb{E}_p[\log p(\mathbf{z})] - \mathbb{E}_p[\log q(\mathbf{z})], \quad (18)$$

where each expectation is taken with respect to the measure  $p$ . The first term on the R.H.S does not depend on the variational parameters. The second term is

$$\begin{aligned} -\mathbb{E}_p[\log q(\mathbf{z})] &= \frac{1}{2} \log |\tilde{\boldsymbol{\Psi}}| + \frac{1}{2} \mathbb{E}_p(\mathbf{z} - \tilde{\boldsymbol{\nu}})^T \tilde{\boldsymbol{\Psi}}^{-1} (\mathbf{z} - \tilde{\boldsymbol{\nu}}) \\ &= \frac{1}{2} \sum_{i=1}^n \log \tilde{\Psi}_{ii} + \frac{1}{\tilde{\Psi}_{ii}} (\Sigma_{ii} + (\mu_i - \tilde{\nu}_i)^2). \end{aligned}$$

This expression is minimized by setting  $\tilde{\nu}_i = \mu_i$  and moreover  $\tilde{\boldsymbol{\nu}} = \boldsymbol{\mu}$ . Differentiating with respect to  $\tilde{\Psi}_{ii}$  and solving at a stationary point, we then have  $\tilde{\Psi}_{ii} = \Sigma_{ii}$ .  $\square$

This next theorem, obtained when minimizing the reverse KL-divergence, is the counterpart to Theorem 3.6.

**Theorem A.4.** *Suppose  $\mathbf{C}$  has constant off-diagonal terms,  $\varepsilon \in [0, 1)$ . When minimizing the reverse KL-divergence, the entropy gap goes to a constant factor in the limit  $n \rightarrow \infty$ , whereas the variance is correctly estimated, that is*

$$\lim_{n \rightarrow \infty} \frac{1}{n} (\mathcal{H}(p) - \mathcal{H}(q)) = \log(1 - \varepsilon), \quad (19)$$

$$\tilde{\Psi}_{ii} = \Sigma_{ii}. \quad (20)$$

*Proof.* The second equality is already stated in Proposition A.3. Note that this result does not depend on the specifics of the example at hand and applies to any covariance matrix.

To obtain the first equality, we start with the shrinkage-linkage decomposition,

$$\mathcal{H}(p) - \mathcal{H}(q) = \frac{1}{2} \log |\mathbf{S}| + \frac{1}{2} \log |\mathbf{C}|. \quad (21)$$

Since there is no shrinkage,  $\log |\mathbf{S}| = 0$ . Next, recall from (14) the expression for  $\log |\mathbf{C}|$ . Dividing  $\log |\mathbf{C}|$  by  $n$  and taking the limit in  $n$ , we obtain the desired result.  $\square$

Naturally, the entropy gap can be arbitrarily large, with  $q$  having a larger entropy than  $p$ , notably as  $\varepsilon$  goes to 1.

For problems where the marginal variances are of interest, rather than the entropy, we would ideally minimize the reverse KL-divergence. Unfortunately, there is usually no efficient way to optimize  $\text{KL}(p||q)$ , due to the difficulty in evaluating expectation values with respect to  $p$ .

## B SOLUTIONS FOR THE BOUNDS ON THE SHRINKAGE AND DELINKAGE TERMS

By exploiting the symmetries that we proved in Section 4.2, we can efficiently compute bounds on the terms  $\log |\mathbf{S}|$  and  $\log |\mathbf{C}|$ ; these are the terms that arise, respectively, from the effects of shrinkage and delinkage.

### B.1 UPPER BOUND ON $\log |\mathbf{S}|$

First we show how to compute the upper bound on  $\log |\mathbf{S}|$ . Recall that to do so, we must solve the optimization problem

$$\max_{\lambda \in \Lambda_R} \sum_{i=1}^n \lambda_i^{-1}. \quad (22)$$

From Lemma 4.2, we know that all the elements of the solution assume the edge values of  $\lambda_1$  or  $\lambda_n$  save for at most one which we denote  $\lambda_k$ . At a high level, we solve the optimization by exhaustively computing the optimal solution for each candidate value of  $k \in \{1, \dots, n\}$ , then choosing the particular value of  $k$  whose solution maximizes the overall objective function.

It remains only to show how to compute the solution for a particular candidate value of  $k$ . Recall the constraints that  $\sum_{i=1}^n \lambda_i = n$  and  $\lambda_1 = R\lambda_n$ . It follows that

$$\lambda_k = n - [(k-1)R + n - k] \lambda_n. \quad (23)$$

Using the constraints to eliminate  $\lambda_1$  and  $\lambda_k$ , we can write the objective function entirely in terms of  $\lambda_n$ . In this way we find

$$\sum_{i=1}^n \lambda_i^{-1} = \frac{1}{n - [R(k-1) + n - k] \lambda_n} + \frac{(k-1)}{R\lambda_n} + \frac{n-k}{\lambda_n}. \quad (24)$$

Crucially, we also need to enforce the boundary conditions  $\lambda_n \leq \lambda_k \leq \lambda_1$ , or equivalently

$$\frac{n}{Rk + n - k} \leq \lambda_n \leq \frac{n}{R(k-1) + n - k + 1}. \quad (25)$$

Note that the simplified objective in (24) for fixed  $k$  is convex in  $\lambda_n$ ; hence the maximizer must lie at one of the boundary values in (25). By computing the objective for each boundary value of  $\lambda_k$ , we find the optimal solution for this

candidate value of  $k$ . Finally, we obtain the overall solution to eq. (22) by considering all  $n$  candidate values of  $k$  and choosing the best one.

### B.2 UPPER BOUND ON $\log |\mathbf{C}|$

Next we show how to compute the upper bound on  $\log |\mathbf{C}|$ . Recall that to do so, we must solve the optimization problem

$$\max_{\lambda \in \Lambda_R} \left[ \sum_{i=1}^n \log \lambda_i \right]. \quad (26)$$

From Lemma 4.3, we know that all eigenvalues other than  $\lambda_1$  and  $\lambda_n$  must have the same value; we denote this value by  $\lambda_k$ . From the constraint  $\sum_{i=1}^n \lambda_i = n$ , it follows that

$$\lambda_k = \frac{n - (1+R)\lambda_n}{n-2}. \quad (27)$$

Again, using the constraints to eliminate  $\lambda_1$  and  $\lambda_k$ , we can write the objective function entirely in terms of  $\lambda_n$ . In this way we find

$$\sum_{i=1}^n \log \lambda_i = (n-2) \log \frac{n - (1+R)\lambda_n}{n-2} + \log R\lambda_n + \log \lambda_n. \quad (28)$$

This objective is concave in  $\lambda_n$ , so we can locate the maximum by setting its derivative with respect to  $\lambda_n$  equal to zero. Some straightforward algebra shows that this derivative vanishes when

$$\lambda_n = \frac{2}{1+R}. \quad (29)$$

Finally we need to check that this solution does not violate the boundary conditions of the problem; in particular, we require that  $\lambda_n \geq \lambda_k \geq R\lambda_n$ , or equivalently that

$$\frac{n}{1+R(n-1)} \leq \lambda_n \leq \frac{n}{n-1+R}. \quad (30)$$

These conditions are always satisfied for  $n \geq 3$ . Hence we obtain an analytical solution for the upper bound on  $\log |\mathbf{C}|$ . Finally, note that while the solution for  $\lambda_n$  does not depend on  $n$ , the optimized objective function does depend on  $n$  through eq. (28).

Algorithm 1 provides an implementation of the above-described method.

## C BOUNDS ON THE AVERAGE VARIANCE SHRINKAGE

We can also derive bounds on the *average* shrinkage in componentwise variance in terms of the problem dimensionality,  $n$ , and the condition number,  $R$ , of the correlation matrix. The average in this case is performed over the different

components of  $\mathbf{z}$ . Recall that the shrinkage in each componentwise variance is given by  $S_{ii} = \Sigma_{ii}/\Psi_{ii}$ . Hence we can also express this bound in terms of the trace of the shrinkage matrix,  $\text{trace}(\mathbf{S})$ .

**Proposition C.1.** *Suppose that the correlation matrix  $\mathbf{C}$  has condition number  $R$ . Then the solution for FG-VI in section 2 satisfies*

$$\min_{\lambda \in \Lambda_R} \sum_{i=1}^n \lambda_i^{-1} \leq \text{trace}(\mathbf{S}) \leq \max_{\lambda \in \Lambda_R} \sum_{i=1}^n \lambda_i^{-1}, \quad (31)$$

where  $\Lambda_R$  is the set defined in section 4.

*Proof.* We showed in the proof of Theorem 3.1 that

$$S_{ii} = \frac{\Sigma_{ii}}{\Psi_{ii}} = C_{ii}^{-1}. \quad (32)$$

It follows that  $\text{trace}(\mathbf{S}) = \text{trace}(\mathbf{C}^{-1}) = \sum_i \lambda_i^{-1}$ , where  $\lambda_1, \dots, \lambda_n$  are the eigenvalues of  $\mathbf{C}$ . The bound then follows from the relaxation from the set  $\mathcal{C}_R$  to the set  $\Lambda_R$  in section 4.  $\square$

### C.1 LOWER BOUND ON $\text{trace}(\mathbf{S})$

The optimization implied by eq. (31) is convex, since both the set  $\Lambda_R$  and the objective function  $\sum_i \lambda_i^{-1}$  are convex. In fact, this bound can be evaluated in closed form by using similar methods as in section 4.2.

**Lemma C.2.** *Let  $\lambda \in \Lambda_R$  be the solution that minimizes the left side of eq. (31). Then  $\lambda_i = \lambda_j$  whenever  $1 < i < j < n$ .*

---

**Algorithm 1:** Upper bounds on  $\log |\mathbf{S}|$  and  $\log |\mathbf{C}|$

---

```

1 Input:  $R, n$ 
2
3 Function  $\text{ObjF}(\lambda_n, k)$ :
4   | return  $(n - k + \frac{k-1}{R}) \frac{1}{\lambda_n} + \frac{1}{n - [R(k-1) + n - k]\lambda_n}$ 
5
6 for  $k$  in  $\{2, \dots, n-1\}$  do
7   |  $\lambda_a \leftarrow \frac{n}{Rk + n - k}$ 
8   |  $\lambda_b \leftarrow \frac{n}{R(k-1) + n - k + 1}$ 
9   |  $F_k \leftarrow \max(\text{ObjF}(\lambda_a), \text{ObjF}(\lambda_b))$ 
10  | if  $(k = 1)$   $F \leftarrow F_k$ 
11  | else  $F \leftarrow \max(F, F_k)$ 
12 end
13  $U_s \leftarrow n \log(F/n)$ 
14
15  $\lambda_n \leftarrow \frac{2}{1+R}$ 
16  $U_c \leftarrow \log \frac{1}{\lambda_n} + \log \frac{1}{R\lambda_n} + (n-2) \log \frac{n-2}{n - (1+R)\lambda_n}$ 
17
18 Return:  $U_s, U_c$ 

```

---

*Proof.* This proof follows the same argument as the proof (by contradiction) for Lemma 4.3. Suppose there exists a solution with intermediate elements  $\lambda_i$  and  $\lambda_j$  that satisfy  $\lambda_1 \geq \lambda_i > \lambda_j \geq \lambda_n$ . Consider the effect on this solution of a perturbation that adds some small amount  $\delta > 0$  to  $\lambda_j$  and subtracts the same amount from  $\lambda_i$ . For sufficiently small  $\delta$ , this perturbation will not leave the set  $\Lambda_R$ ; however, it will *diminish* the separation of  $\lambda_i$  from  $\lambda_j$ . As a result the objective  $\sum_k (1/\lambda_k)$  experiences a change

$$g(\delta) = \frac{1}{\lambda_i - \delta} - \frac{1}{\lambda_i} + \frac{1}{\lambda_j + \delta} - \frac{1}{\lambda_j}. \quad (33)$$

Evaluating the derivative, we find  $g'(0) = \lambda_i^{-2} - \lambda_j^{-2} < 0$ , so that the objective function is decreased for some  $\delta > 0$ . As before this yields a contradiction, because any solution must be maximal, and hence stationary (i.e.,  $g'(0) = 0$ ), with respect to small perturbations.  $\square$

With the above lemma, the  $n$ -dimensional optimization over  $\Lambda_R$  can be reduced to a one-dimensional optimization that can be solved in closed form. The methods are identical to those in the previous appendix.

First we rewrite the constraint,  $\lambda_n \leq \lambda_k \leq R\lambda_n$ , as

$$\frac{n}{R(n-1)+1} \leq \lambda_n \leq \frac{n}{R+n-1}. \quad (34)$$

Since the minimization problem is convex, a minima can be found at a stationary point of the objective function

$$\sum_{i=1}^n \frac{1}{\lambda_i} = \frac{(n-2)^2}{n - (1+R)\lambda_n} + \frac{1}{\lambda_n} + \frac{1}{R\lambda_n}, \quad (35)$$

which now only depends on  $\lambda_n$ . Differentiating and setting to 0, we obtain the root-finding problem,

$$[R(n-2)^2 - (1+R)^2] \lambda_n^2 + 2n(1+R)\lambda_n - n^2 = 0, \quad (36)$$

which can be solved exactly. It remains to check whether the roots violate the constraints in (34), and pick the non-offending root which maximizes the objective in (35). If both roots violate the constraints then, by convexity of the problem, the solution must lie at one of the boundary terms in (34).

### C.2 UPPER BOUND ON $\text{TRACE}(\mathbf{S})$

A similar approach gives us an upper bound on  $\text{trace}(\mathbf{S})$ . In fact, we have already solved the problem of maximizing the right side of (31) when upper-bounding  $\log |\mathbf{S}|$ . It remains to apply the same strategy.

## D TIGHTER UPPER BOUND ON ENTROPY GAP

In Proposition 4.1 we derived separate upper bounds on the individual terms  $\log |\mathbf{S}|$  and  $\log |\mathbf{C}|$ . One upper bound on

the entropy gap (or equivalently, on  $\text{KL}(q, p)$ ) is obtained simply by adding these separate bounds. However, a tighter upper bound is obtained by replacing the separate optimizations in Proposition 4.1 by a single joint optimization:

$$\text{KL}(q, p) \leq \frac{1}{2} \max_{\lambda \in \Lambda_R} \left[ n \log \frac{1}{n} \sum_{i=1}^n \lambda_i^{-1} + \sum_{i=1}^n \log \lambda_i \right]. \quad (37)$$

In this appendix we sketch how to solve this optimization and evaluate this bound in closed form. The first step is to make the change of variables,

$$\omega_i = \frac{\lambda_i^{-1}}{\sum_{j=1}^n \lambda_j^{-1}}, \quad (38)$$

and to translate the domain of optimization accordingly. Under this change of variables, the original domain  $\Lambda_R$  in section 4 is mapped onto the set

$$\Omega_R = \left\{ \omega \in \mathbb{R}_+^n \mid \omega_n \geq \dots \geq \omega_1 = \frac{1}{R} \omega_n, \sum_{i=1}^n \omega_i = 1 \right\}. \quad (39)$$

Likewise, a little algebra shows that the optimization in eq. (37) is equivalent to the following:

$$\text{KL}(q, p) \leq \frac{1}{2} \max_{\omega \in \Omega_R} \left[ \sum_{i=1}^n \log \frac{1}{\omega_i} - n \log n \right]. \quad (40)$$

Now we can make a similar argument as in the proof of Lemma 4.2 to simplify this optimization.

**Lemma D.1.** *Let  $\omega \in \Omega_R$  be the solution that maximizes the right side of eq. (40). Then at most one  $\omega_i$  is not equal to either  $\omega_1$  or  $\omega_n$ .*

*Proof.* We prove the lemma by contradiction. Suppose there exists a solution with intermediate elements  $\omega_i$  and  $\omega_j$  that satisfy  $\omega_n > \omega_i > \omega_j > \omega_1$ . Consider the effect on this solution of a perturbation that adds some small amount  $\delta > 0$  to  $\omega_i$  and subtracts the same amount from  $\omega_j$ . Note that for sufficiently small  $\delta$ , this perturbation will not leave the set  $\Omega_R$ ; however, it will *expand* the separation of  $\omega_i$  from  $\omega_j$ . As a result the objective in eq. (40) changes by an amount

$$f(\delta) = \frac{1}{2} \left[ \log \frac{1}{\omega_i + \delta} - \log \frac{1}{\omega_i} + \log \frac{1}{\omega_j - \delta} - \log \frac{1}{\omega_j} \right]. \quad (41)$$

Next we evaluate the derivative  $f'(\delta)$  at  $\delta = 0$ ; doing so we find  $f'(0) = \omega_j^{-1} - \omega_i^{-1} > 0$ , so that the objective is increased for some  $\delta > 0$ . But this yields a contradiction, because any solution must be maximal, and hence stationary (i.e.,  $f'(0) = 0$ ), with respect to small perturbations.  $\square$

With the above lemma, we can reduce the  $n$ -dimensional optimization over  $\Omega_R$  to a one-dimensional optimization

that can be solved in closed form; the methods are identical to those in the previous appendix.

In details, let  $\omega_k$  be the one variable which (potentially) does not go to  $\omega_1$  or  $\omega_n$ . Given  $\sum_{i=1}^n \omega_i = 1$ ,

$$\omega_k = 1 - (k - 1 + R(n - k))\omega_1. \quad (42)$$

The objective is then

$$\sum_{i=1}^n \log \frac{1}{\omega_i} = -(k - 1) \log \omega_1 - (n - k) \log R\omega_1 - \log(1 - [k - 1 + R(n - k)\omega_1]). \quad (43)$$

Since we are trying to maximize a convex function, the solution does not lie at a stationary point, rather at a boundary set by the constraint,  $\omega_1 \leq \omega_k \leq \omega_n$ , or equivalently

$$\frac{1}{k - 1 + R(n - k + 1)} \leq \omega_1 \leq \frac{1}{k + R(n - k)}. \quad (44)$$

It remains to test each candidate boundary for each choice of  $k$  to obtain a maximizer.