Appendix A
Derivation of A

Recall that $Z(x,\beta) := \sum_{\hat{x}} p_\beta(\hat{x}) e^{-\beta d(x,\hat{x})}$, therefore

$$\frac{\partial Z(x,\beta)}{\partial p_\beta(\hat{x})} = e^{-\beta d(x,\hat{x})} \ . \tag{16}$$

For $\hat{x}' \neq \hat{x}$ we have

$$\frac{\partial}{\partial p_\beta(\hat{x}')} \left(\frac{p_\beta(\hat{x})}{Z(x,\beta)} \right) = - \frac{p_\beta(\hat{x}) e^{-\beta d(x,\hat{x}')}}{Z(x,\beta)^2} \ , \tag{17}$$

and thus

$$\frac{\partial F\big(p_\beta(\hat{x})\big)}{\partial p_\beta(\hat{x}')} = p_\beta(\hat{x}) \sum_x p(x) \frac{e^{-\beta\left(d(x,\hat{x})+d(x,\hat{x}')\right)}}{Z(x,\beta)^2} \ . \tag{18}$$

In contrast,

$$\begin{aligned} \frac{\partial}{\partial p_\beta(\hat{x})} \left(\frac{p_\beta(\hat{x})}{Z(x,\beta)} \right) &= \frac{Z(x,\beta) - p_\beta(\hat{x}) e^{-\beta d(x,\hat{x})}}{Z(x,\beta)^2} \\ &= \frac{1}{Z(x,\beta)} - \frac{p_\beta(\hat{x}) e^{-\beta d(x,\hat{x})}}{Z(x,\beta)^2} \ , \end{aligned} \tag{19}$$

and then

$$\begin{aligned} \frac{\partial F\big(p_\beta(\hat{x})\big)}{\partial p_\beta(\hat{x})} &= 1 - \sum_x p(x) \left(\tfrac{e^{-\beta d(x,\hat{x})}}{Z(x,\beta)} - \tfrac{p_\beta(\hat{x}) e^{-2\beta d(x,\hat{x})}}{Z(x,\beta)^2} \right) \\ &= 1 - \sum_x \tfrac{p(x) e^{-\beta d(x,\hat{x})}}{Z(x,\beta)} + p_\beta(\hat{x}) \sum_x p(x) \tfrac{e^{-2\beta d(x,\hat{x})}}{Z(x,\beta)^2} \ . \end{aligned} \tag{20}$$

When $p_\beta(\hat{x}) > 0$ we have by (2), applying Bayes' law,

$$p_\beta(x|\hat{x}) = \frac{p(x)}{p_\beta(\hat{x})} p(\hat{x}|x) = \frac{p(x) e^{-\beta d(x,\hat{x})}}{Z(x,\beta)} \ , \tag{21}$$

and therefore $\sum_x \frac{p(x) e^{-\beta d(x,\hat{x})}}{Z(x,\beta)} = 1$. Note, however, that the right hand side of (21) is well defined even when $p_\beta(\hat{x}) = 0$, and since we are interested only in the right derivative when $p_\beta(\hat{x}) = 0$ (as it is in the boundary of $\Delta\hat{\mathcal{X}}$), we can refer in that case to the limit, which also satisfies $\sum_x \frac{p(x) e^{-\beta d(x,\hat{x})}}{Z(x,\beta)} = 1$. Consequently, we have from (20)

$$\frac{\partial F\big(p_\beta(\hat{x})\big)}{\partial p_\beta(\hat{x})} = p_\beta(\hat{x}) \sum_x p(x) \frac{e^{-2\beta d(x,\hat{x})}}{Z(x,\beta)^2} \ , \tag{22}$$

which together with (18) gives the result in (7).

The formula in (6) is a straightforward result of (7) and (21). Although simpler, it is not defined when $p_\beta(\hat{x}) = 0$.

Appendix B
Proof of Theorem 1

Proof of Lemma 2: If $p_\beta(\hat{x}) = 0$ then by (7) the corresponding column of A is zero and thus $A\mathbf{e}_{\hat{x}} = 0$. Conversely, if $A\mathbf{e}_{\hat{x}} = 0$, then by (7) we have particularly

$$p_\beta(\hat{x}) \sum_x p(x) \frac{e^{-2\beta d(x,\hat{x})}}{Z(x,\beta)^2} = 0 \ . \tag{23}$$

Therefore, either $p_\beta(\hat{x}) = 0$ or $d(x,\hat{x}) = \infty$ for all x. But the latter implies by (2) that $p_\beta(\hat{x}|x) = 0$ for all x, and so anyway $p_\beta(\hat{x}) = 0$. ■

Proof of Proposition 3: We prove the proposition by induction on r.

The case $r = 1$ is trivial. Let $r = 2$ and assume by contradiction that $\mathbf{e}_{\hat{x}_1} \notin \ker A$. This means that also $\mathbf{e}_{\hat{x}_2} \notin \ker A$, otherwise $v - v_{\hat{x}_2}\mathbf{e}_{\hat{x}_2} \in \ker A$, which would imply that $\mathbf{e}_{\hat{x}_1} \in \ker A$. Therefore, according to Lemma 2, both $p_\beta(\hat{x}_1), p_\beta(\hat{x}_2) > 0$.

Now, since $v \in \ker A$, we have by (7) for all $\hat{x}$

$$\begin{aligned} 0 &= \sum_{\hat{x}'} p_\beta(\hat{x}') \sum_x p(x) \frac{e^{-\beta\left(d(x,\hat{x})+d(x,\hat{x}')\right)}}{Z(x,\beta)^2} v_{\hat{x}'} && (24) \\ &= \sum_x p(x) \frac{e^{-\beta d(x,\hat{x})}}{Z(x,\beta)} \Big(p_\beta(\hat{x}_1|x) v_{\hat{x}_1} + p_\beta(\hat{x}_2|x) v_{\hat{x}_2} \Big) \ , && (25) \end{aligned}$$

where the second equality follows from (2). Averaging over $p_\beta(\hat{x})$ gives $p_\beta(\hat{x}_1)v_{\hat{x}_1} + p_\beta(\hat{x}_2)v_{\hat{x}_2} = 0$, and thus $v_{\hat{x}_1} = -\frac{p_\beta(\hat{x}_2)}{p_\beta(\hat{x}_1)} v_{\hat{x}_2}$. Plugging this result back in (25) and dividing by $p_\beta(\hat{x}_2)\, v_{\hat{x}_2} \neq 0$ we get for all $\hat{x}$

$$\sum_x p(x) \frac{e^{-\beta d(x,\hat{x})}}{Z(x,\beta)^2} \left(e^{-\beta d(x,\hat{x}_2)} - e^{-\beta d(x,\hat{x}_1)} \right) = 0 \ , \tag{26}$$

where we used again (2). Substituting $\hat{x} = \hat{x}_1, \hat{x}_2$ in the last equation we have

$$\begin{aligned} \sum_x \tfrac{p(x)}{Z(x,\beta)^2} \left(e^{-\beta d(x,\hat{x}_1)} e^{-\beta d(x,\hat{x}_2)} - e^{-2\beta d(x,\hat{x}_1)} \right) = 0 && (27) \\ \sum_x \tfrac{p(x)}{Z(x,\beta)^2} \left(e^{-2\beta d(x,\hat{x}_2)} - e^{-\beta d(x,\hat{x}_1)} e^{-\beta d(x,\hat{x}_2)} \right) = 0 && (28) \end{aligned}$$

and subtracting (27) from (28) gives

$$\sum_x \frac{p(x)}{Z(x,\beta)^2} \left(e^{-\beta d(x,\hat{x}_1)} - e^{-\beta d(x,\hat{x}_2)} \right)^2 = 0 \ . \tag{29}$$

Therefore, for all x we must have $d(x,\hat{x}_1) = d(x,\hat{x}_2)$, contradicting our assumption on the non-degeneracy of d. Consequently, $\mathbf{e}_{\hat{x}_1} \in \ker A$, and thus $v - v_{\hat{x}_2}\mathbf{e}_{\hat{x}_2} \in \ker A$, implying that also $\mathbf{e}_{\hat{x}_1} \in \ker A$.

Finally, let $r \geq 3$ and assume the proposition holds for all $1 \leq r' < r$. If there exists $1 \leq i \leq r$ such that $\mathbf{e}_{\hat{x}_i} \in \ker A$, then $u = v - v_{\hat{x}_i}\mathbf{e}_{\hat{x}_i} \in \ker A$. However, u has exactly $r-1 < r$ nonzero coordinates, namely $\hat{x}_j$ for $1 \leq j \leq r,\ j \neq i$, and therefore by the induction hypothesis all the corresponding $\mathbf{e}_{\hat{x}_j}$ also belong to $\ker A$. Together with $\mathbf{e}_{\hat{x}_i}$ this completes the induction step.

Conversely, assume by contradiction that $\mathbf{e}_{\hat{x}_i} \notin \ker A$ for all $1 \leq i \leq r$, then by Lemma 2 we have $p_\beta(\hat{x}_i) > 0$ for all $\hat{x}_i$. This implies that $A_{\hat{x}\hat{x}_i} > 0$ for all $\hat{x}$ and $\hat{x}_i$, otherwise by (7) we would get for some $\hat{x}$ and $\hat{x}_i$ that $d(x,\hat{x}) + d(x,\hat{x}_i) = \infty$ for all x, contradicting our assumption on the finiteness of d. In particular, this means that $\sum_{i=2}^{r} A_{\hat{x}\hat{x}_i} > 0$.

Now, since $v \in \ker A$ we have $\sum_{i=1}^{r} A_{\hat{x}\hat{x}_i} v_{\hat{x}_i} = 0$ for all $\hat{x}$, and thus

$$0 = A_{\hat{x}\hat{x}_1} v_{\hat{x}_1} + \sum_{i=2}^{r} A_{\hat{x}\hat{x}_i} v_{\hat{x}_i} \tag{30}$$

$$= \sum_{i=2}^{r} \frac{A_{\hat{x}\hat{x}_i} A_{\hat{x}\hat{x}_1} v_{\hat{x}_1}}{\sum_{j=2}^{r} A_{\hat{x}\hat{x}_j}} + \sum_{i=2}^{r} A_{\hat{x},\hat{x}_i} v_{\hat{x}_i} \tag{31}$$

$$= \sum_{i=2}^{r} A_{\hat{x}\hat{x}_i} \left(\frac{A_{\hat{x}\hat{x}_1} v_{\hat{x}_1}}{\sum_{j=2}^{r} A_{\hat{x}\hat{x}_j}} + v_{\hat{x}_i} \right) . \tag{32}$$

Define the vector $u \in \mathbb{R}^m$ such that $u_{\hat{x}_i} = \frac{A_{\hat{x}\hat{x}_1} v_{\hat{x}_1}}{\sum_{j=2}^{r} A_{\hat{x}\hat{x}_j}} + v_{\hat{x}_i}$ for all $2 \le i \le r$, and all its other coordinates are 0. By (32) $u \in \ker A$ and it has at most $r-1 < r$ nonzero coordinates. If there exists $2 \le i \le r$ such that $u_{\hat{x}_i} \neq 0$ then by the induction hypothesis we would have $\mathbf{e}_{\hat{x}_i} \in \ker A$, contradicting our assumption. Therefore, for all $2 \le i \le r$ we must have

$$v_{\hat{x}_i} = -\frac{A_{\hat{x}\hat{x}_1} v_{\hat{x}_1}}{\sum_{j=2}^{r} A_{\hat{x}\hat{x}_j}} . \tag{33}$$

In particular this implies that $\operatorname{sgn} v_{\hat{x}_2} = \operatorname{sgn} v_{\hat{x}_3} = -\operatorname{sgn} v_{\hat{x}_1}$. Finally, we can perform the same analysis starting at (30) by setting aside $v_{\hat{x}_2}$ instead of $v_{\hat{x}_1}$, concluding with $\operatorname{sgn} v_{\hat{x}_1} = \operatorname{sgn} v_{\hat{x}_3} = -\operatorname{sgn} v_{\hat{x}_2}$. Together with the previous result, this means that $\operatorname{sgn} v_{\hat{x}_i} = 0$ for $i = 1, 2, 3$, or equivalently that $v_{\hat{x}_i} = 0$, contradicting our initial assumption and completing the induction step. ■

APPENDIX C
PROOF OF THEOREM 4

Proof: First, we deal with the case in which all $p_\beta(\hat{x}) > 0$. Note from (7) that A can be written as the product of three matrices,

$$A = BB^\intercal C , \tag{34}$$

where $B_{\hat{x}x} = \frac{p(x)^{1/2}}{Z(x,\beta)} e^{-\beta d(x,\hat{x})}$ and C is a diagonal matrix with $p_\beta(\hat{x})$ in its diagonal. Therefore, we have

$$C^{1/2} A C^{-1/2} = C^{1/2} BB^\intercal C^{1/2} = (C^{1/2}B)(C^{1/2}B)^\intercal , \tag{35}$$

meaning that A is similar to a real Gram matrix, and thus diagonalizable with non-negative eigenvalues [27] .

Second, assume that $p_\beta(\hat{x}_i) = 0$ for $i = 1, \dots, r$ – the first $r \ge 1$ coordinates – and $p_\beta(\hat{x}) > 0$ elsewhere. Let $\hat{X}' = \operatorname{supp} p_\beta$ and denote by p'_β and A' the solutions and matrix corresponding to the RD problem restricted to $\hat{X}'$. Note that $Z(x, \beta)$ depends only on the support, and thus

$$A = \begin{pmatrix} 0 & \cdots \\ 0 & A' \end{pmatrix} . \tag{36}$$

Since $p'_\beta(\hat{x}) > 0$ for all $\hat{x} \in \hat{X}'$, there exist, by the first part of the proof, an invertible matrix P and a non-negative diagonal matrix Λ such that $P^{-1}A'P = \Lambda$. Moreover, from Theorem 1 we have $\dim \ker A' = 0$, hence none of the values in the diagonal of Λ (that is, the eigenvalues of A') is 0.

Now, we have

$$\begin{aligned} &\begin{pmatrix} I_r & 0 \\ 0 & P^{-1} \end{pmatrix} A \begin{pmatrix} I_r & 0 \\ 0 & P \end{pmatrix} \\ &\quad = \begin{pmatrix} I_r & 0 \\ 0 & P^{-1} \end{pmatrix} \begin{pmatrix} 0 & \cdots \\ 0 & A' \end{pmatrix} \begin{pmatrix} I_r & 0 \\ 0 & P \end{pmatrix} \\ &\quad = \begin{pmatrix} 0 & \cdots \\ 0 & P^{-1}A'P \end{pmatrix} = \begin{pmatrix} 0 & \cdots \\ 0 & \Lambda \end{pmatrix} , \end{aligned} \tag{37}$$

where I_r is the $r \times r$ identity matrix. This means that A is similar to an upper-triangular matrix, and thus all its eigenvalues appear in that matrix diagonal, repeated according to their respective algebraic multiplicities. Since Λ has no zeroes in its diagonal, we conclude that the algebraic multiplicity of the eigenvalue 0 of A is exactly r. However, according to Theorem 1 we have $\dim \ker A = r$, or equivalently, that the geometric multiplicity of the eigenvalue 0 of A is also r.

Finally, note that the standard basis row vector $\mathbf{e}^\intercal_{\hat{x}_{r+i}}$ for $i \ge 1$ is a left eigenvector of the matrix in (37), associated with the eigenvalue Λ_{ii}. Therefore, also for all nonzero eigenvalues of A, the algebraic multiplicity must equal the geometric multiplicity. Consequently the matrix A is diagonalizable with real eigenvalues. ■

APPENDIX D
PROOF OF THEOREM 5

Proof: Denote by $\tilde{\delta p}_k$ the deviation vector $p_k - p_\beta$ of the k-th iterate from the fixed point p_β, $\tilde{\delta p}_k(\hat{x}) = p_k(\hat{x}) - p_\beta(\hat{x})$ for its $\hat{x}$-indexed entry. For convenience, we use the L^1 norm in the sequel, denoted $\|\cdot\|$. The expansion of AB around p_β is [30]

$$AB(p_\beta + \tilde{\delta p}_k) - p_\beta = \nabla AB\big|_{p_\beta} \tilde{\delta p}_k + O(\|\tilde{\delta p}_k\|^2). \tag{38}$$

That is, to first order, a single AB iteration amounts to an application of the linear operator $\nabla AB\big|_{p_\beta(\hat{x})} = I - A^\intercal$ to the deviation. Write $B(0, r)$ for the ball of radius r around the origin with respect to L^1. Then,

$$\tilde{\delta p}_{k+1} \in (I - A^\intercal)\tilde{\delta p}_k + B(0, \tilde{c}\|\tilde{\delta p}_k\|^2), \tag{39}$$

where $\tilde{c} > 0$ is a constant bounding the expansion's remainder. By Theorem 4, A is diagonalizable, and so $I - A^\intercal = P\Lambda P^{-1}$ with Λ diagonal. Multiplying (39) by P^{-1},

$$P^{-1}\tilde{\delta p}_{k+1} \in P^{-1}\left(P\Lambda P^{-1}\right)\tilde{\delta p}_k + P^{-1}B(0, \tilde{c}\|\tilde{\delta p}_k\|^2). \tag{40}$$

Denote $\|\cdot\|_{op}$ for the operator norm with respect to L^1. By its definition, $P^{-1}B(0, r) \subset B(0, \|P^{-1}\|_{op}\, r)$. Thus, exchanging coordinates $\delta p_k := P^{-1}\tilde{\delta p}_k$ to a basis of eigenvectors,

$$\delta p_{k+1} \in \Lambda \delta p_k + B(0, c\|\delta p_k\|^2), \tag{41}$$

for $c := \tilde{c} \cdot \|P\|^2_{op} \cdot \|P^{-1}\|_{op}$.

Denote by $\lambda_{max} = \lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ the eigenvalues of $I - A^\intercal$. As noted at Section IV, they are contained in $[0, 1)$ by our assumptions. Denote the i-th coordinate of δp_k with respect to this basis by $\delta p_k^{(i)}$, $i = 1, \dots, |\hat{\mathcal{X}}|$. For exposition's simplicity, suppose that λ_1 is a simple eigenvalue, $\lambda_1 > \lambda_2$; the proof is similar otherwise[3].

Let $0 < a < \frac{1}{-\log \lambda_1}$. An upper bound for convergence is immediate, when $\lambda_1 < 1$. Choose $\mu := \exp\left((\frac{1}{\log \lambda_1} - a)^{-1} \right)$. It satisfies $\frac{1}{-\log \mu} = \frac{1}{-\log \lambda_1} + a$, and $\lambda_1 < \mu < 1$. Then whenever $\|\delta p_k\| \le \frac{1}{c}(\mu - \lambda_1)$ we have

$$\|\delta p_{k+1}\| \overset{(41)}{\le} \lambda_1 \|\delta p_k\| + c\|\delta p_k\|^2 \le \mu \|\delta p_k\| \,. \tag{42}$$

Since $\|\tilde{\delta p_k}\| \le \|P\|_{op} \cdot \|\delta p_k\|$, this holds whenever

$$\|\tilde{\delta p_k}\| \le \delta_1 := \tfrac{\|P\|_{op}}{c}(\mu - \lambda_1). \tag{43}$$

Therefore, at most

$$k \le \frac{-\log \varepsilon + \log(\|P^{-1}\tilde{\delta p_0}\| \cdot \|P\|_{op})}{-\log \mu} \tag{44}$$

iterations are then required for ε-convergence of p_k. To capture the asymptotic convergence rate we divide by $-\log \varepsilon$ and take the limit to obtain

$$\lim_{\varepsilon \to 0^+} \frac{k}{-\log \varepsilon} \le \lim_{\varepsilon \to 0^+} \frac{1 + \frac{\log(\|P^{-1}\tilde{\delta p_0}\| \cdot \|P\|_{op})}{-\log \varepsilon}}{-\log \mu} = \\ = \frac{1}{-\log \mu} = \frac{1}{-\log \lambda_1} + a. \tag{45}$$

For a lower bound, choose $\eta := \exp\left((\frac{1}{\log \lambda_1} + a)^{-1} \right)$. It satisfies $\frac{1}{-\log \eta} = \frac{1}{-\log \lambda_1} - a > 0$, and thus $0 < \eta < \lambda_1$. Define,

$$\rho(\delta p) := \frac{|\delta p^{(1)}|}{\|\delta p\|} \tag{46}$$

when $\delta p^{(1)} \ne 0$, $\rho_k := \rho(\delta p_k)$. We proceed by assuming

$$|\delta p_k^{(1)}| \ge \rho_0 \cdot \|\delta p_k\| > 0 \tag{47}$$

for all $k \ge 0$. That is, the relative weight of the first components cannot decrease beyond its initial value at $k = 0$. This shall be justified in the sequel. From (41),

$$|\delta p_{k+1}^{(1)}| \ge \lambda_1 |\delta p_k^{(1)}| - c\|\delta p_k\|^2 \overset{(47)}{\ge} \lambda_1 |\delta p_k^{(1)}| - c\tfrac{1}{\rho_0^2} |\delta p_k^{(1)}|^2 = \\ = |\delta p_k^{(1)}| \left[\lambda_1 - \tfrac{c}{\rho_0^2} |\delta p_k^{(1)}| \right]. \tag{48}$$

Thus, if $|\delta p_k^{(1)}| \le \frac{\rho_0^2}{c}(\lambda_1 - \eta)$ then $|\delta p_{k+1}^{(1)}| \ge \eta |\delta p_k^{(1)}|$. If the above were to hold for all $k \ge 0$, then we obtain a lower bound

$$|\delta p_k^{(1)}| \ge \eta^k \, |\delta p_0^{(1)}|. \tag{49}$$

Since $|\delta p_k^{(1)}| \le \|\delta p_k\| \le \|P^{-1}\|_{op} \cdot \|\tilde{\delta p_k}\|$, the condition $|\delta p_k^{(1)}| \le \frac{\rho_0^2}{c}(\lambda_1 - \eta)$ can be replaced by the stricter

$$\|\tilde{\delta p_k}\| \le \delta_2 := \tfrac{\rho_0^2}{c\|P^{-1}\|_{op}}(\lambda_1 - \eta). \tag{50}$$

[3]If λ_{max} is of multiplicity > 1, then take $\delta p_k^{(1)}$ to be a non-zero component along some normalized λ_{max}-eigenvector, and discard the other coordinates in the λ_{max}-eigenspace. The proof follows with minor modifications.

Since $\|\delta p_k\| \ge |\delta p_k^{(1)}|$ and $|\delta p_0^{(1)}| \ge \rho_0 \|\delta p_0\|$ by assumption, then (49) implies

$$\|P^{-1}\|_{op} \cdot \|\tilde{\delta p_k}\| \ge \|\delta p_k\| \ge \eta^k \cdot \rho_0 \|\delta p_0\| = \eta^k \cdot \rho_0 \|P^{-1}\tilde{\delta p_0}\|. \tag{51}$$

Thus, at least

$$k \ge \frac{-\log \varepsilon + \log(\rho_0 \frac{\|P^{-1}\tilde{\delta p_0}\|}{\|P^{-1}\|_{op}})}{-\log \eta} \tag{52}$$

iterations are required for ε-convergence of p_k. In a manner similar to before,

$$\lim_{\varepsilon \to 0^+} \frac{k}{-\log \varepsilon} \ge \lim_{\varepsilon \to 0^+} \frac{1 + \frac{\log(\frac{\rho_0 \|P^{-1}\tilde{\delta p_0}\|}{\|P^{-1}\|_{op}})}{-\log \varepsilon}}{-\log \eta} = \\ = \frac{1}{-\log \eta} = \frac{1}{-\log \lambda_1} - a. \tag{53}$$

Next, we prove assumption (47) by induction. That is, that the relative weight of the first component cannot decrease beyond ρ_0. For $k = 0$ this is the definition of ρ_0. Assuming it holds for k, we shall prove that it holds for $k+1$. i.e., we shall prove

$$|\delta p_{k+1}^{(1)}| \ge \rho_0 \cdot \|\delta p_{k+1}\|. \tag{54}$$

For the right-hand side of (54),

$$\|\delta p_{k+1}\| = |\delta p_{k+1}^{(1)}| + \sum_{i=2}^{n} |\delta p_{k+1}^{(i)}| \overset{(41)}{\le} \\ \le \lambda_1 |\delta p_k^{(1)}| + \lambda_2 \sum_{i=2}^{n} |\delta p_k^{(i)}| + c\|\delta p_k\|^2 = \\ = \lambda_1 |\delta p_k^{(1)}| + \lambda_2 \left(\|\delta p_k\| - |\delta p_k^{(1)}| \right) + c\|\delta p_k\|^2 = \\ = (\lambda_1 - \lambda_2)|\delta p_k^{(1)}| + \lambda_2 \|\delta p_k\| + c\|\delta p_k\|^2 \tag{55}$$

Thus, using the lower bound (48) on $|\delta p_{k+1}^{(1)}|$, to prove (54) it suffices to show that,

$$|\delta p_k^{(1)}| \left[\lambda_1 - \tfrac{c}{\rho_0^2} |\delta p_k^{(1)}| \right] \ge \\ \ge \rho_0 \left\{ (\lambda_1 - \lambda_2)|\delta p_k^{(1)}| + \lambda_2 \|\delta p_k\| + c\|\delta p_k\|^2 \right\}. \tag{56}$$

By the induction assumption (47), $\frac{1}{\rho_0}|\delta p_k^{(1)}| \ge \|\delta p_k\|$. So, the latter is implied by the stricter,

$$\lambda_1 - \tfrac{c}{\rho_0^2} |\delta p_k^{(1)}| \ge \rho_0 \left\{ (\lambda_1 - \lambda_2) + \tfrac{\lambda_2}{\rho_0} + \tfrac{c}{\rho_0^2} |\delta p_k^{(1)}| \right\}. \tag{57}$$

This reduces to,

$$|\delta p_k^{(1)}| \le \frac{\rho_0^2(1 - \rho_0)(\lambda_1 - \lambda_2)}{c(1 + \rho_0)}. \tag{58}$$

Using $1 + \rho_0 \le 2$, and $|\delta p_k^{(1)}| \le \|P^{-1}\|_{op} \cdot \|\tilde{\delta p_k}\|$ again, this is implied by the stricter,

$$\|\tilde{\delta p_k}\| \le \delta_3 := \frac{\rho_0^2(1 - \rho_0)(\lambda_1 - \lambda_2)}{2c\|P^{-1}\|_{op}}. \tag{59}$$

This guarantees that the induction step (54) holds.

To complete the proof, consider δ_2 (50) and δ_3 (59) as functions of ρ_0, $\delta_2 = \delta_2(\rho_0), \delta_3 = \delta_3(\rho_0)$. Let $B(\delta)$ be the ball of radius δ around p_β, and

$$\tilde{B}(\delta) := \Big\{ p_0 \in B(\delta) : \|p_0 - p_\beta\| \leq \\ \leq \min\{\delta_1, \delta_2\left(\rho(p_0 - p_\beta)\right), \delta_3\left(\rho(p_0 - p_\beta)\right)\} \Big\}. \quad (60)$$

That is, $\tilde{B}(\delta)$ consists of the initial conditions in $B(\delta)$ whose relative weight of the first coordinate satisfies the constraints imposed by δ_1, δ_2 and δ_3.

Clearly, $\tilde{B}(\delta) \subset B(\delta)$. However, notice that δ_1 is strictly positive, as are $\delta_2(\rho_0)$ and $\delta_3(\rho_0)$ for $0 < \rho_0 < 1$. They are polynomials of zeroth, second and third order in ρ_0. Given $\delta > 0$ small enough, one can solve the equations $\delta \leq \delta_2(\rho_0)$ and $\delta \leq \delta_3(\rho_0)$ in ρ_0, to obtain a maximal interval $[\rho_0^-(\delta), \rho_0^+(\delta)]$ for which all the initial conditions $p_0 \in B(\delta)$ with $\rho(p_0) \in [\rho_0^-(\delta), \rho_0^+(\delta)]$ are also in $\tilde{B}(\delta)$. Notice that the intervals $[\rho_0^-(\delta), \rho_0^+(\delta)] \subset [0, 1]$ increase as $\delta > 0$ becomes smaller, and $\lim_{\delta \to 0^+}[\rho_0^-(\delta), \rho_0^+(\delta)] = (0, 1)$. Thus

$$\lim_{\delta \to 0} \frac{\operatorname{vol} \tilde{B}(\delta)}{\operatorname{vol} B(\delta)} = 1, \quad (61)$$

for $\operatorname{vol} S$ the volume of a set S. Since $\tilde{B}(\delta)$ are initial conditions for which the lower and upper bounds hold with the pre-specified accuracy a, this completes the proof. ■