Perron-Frobenius Theory and Applications

M.C. Spruill

January 29, 2007

1 Definitions and basic properties

Let A be an $n \times n$ real matrix. $A \geq 0$ means all entries non-negative, $A > 0$ means all non-negative and at least one positive, and $A >> 0$ means all entries strictly positive. Let $A \geq 0$ and

$$\Lambda = \{\lambda \text{ real } : \text{for some } x \in R^n, x > 0 \text{ and } x'1 = 1 \text{ one has } Ax \geq \lambda x\}.$$

Clearly $\Lambda \neq \emptyset$ since $0 \in \Lambda$. Let

$$\lambda_0 = \sup\{\lambda : \lambda \in \Lambda\}.$$

Lemma 1 *λ_0 is finite, and if $A >> 0$, then λ_0 is positive.*

Proof: Suppose there is a sequence $\lambda_n \to \infty$ of points in Λ. Then there is a sequence of unit vectors x_n which, lying in a compact set, has a subsequence $x_{n'}$ and an x^* to which they converge such that for all n'

$$Ax_{n'} \geq \lambda_{n'} x_{n'}.$$

Since $x_{n'} > 0$ and $x'_{n'}1 = 1$ it follows that $x^* > 0$ and ${x^*}'1 = 1$. Therefore, there is at least one non-zero component, say x^*_1, of x^*. So if N is such that $n' > N$ entails $x_{n'1} > x^*_1/2$ one has

$$\lambda_{n'} x^*_1/2 \leq \sum_{j=1}^{n} a_{1j} x_{n'j} \leq \max\{a_{1j}\} x'_{n'} 1 = \max\{a_{1j}\}$$

for all $n' > N$. This entails at least one element of A being $+\infty$. This is impossible, so λ_0 is finite.

Suppose that $A >> 0$. Let $\delta \in (0, \min\{a_{ij}\})$. Then for $x > 0$, $(A - \delta I)x >> 0$. Therefore $\delta \in \Lambda$ and λ_0 is positive. $\square$

Theorem 1 *If $A >> 0$ then*

- i) there is a vector $x_0 >> 0$ such that $Ax_0 = \lambda_0 x_0$.
- ii) If $\lambda \neq \lambda_0$ is any other eigenvalue of A, then $|\lambda| < \lambda_0$.
- iii) The right eigenvectors of A corresponding to the eigenvalue λ_0 form a one dimensional subspace of R^n.

Proof:

i) - Consulting the argument above, there is an x^* and sequences $\lambda_n \to \lambda_0$ and $x_n \to x^*$ such that

$$Ax^* \geq \lambda_0 x^*.$$

Suppose that strict inequality holds. Then $Ax^* - \lambda_0 x^* > 0$ so that $A(Ax^* - \lambda_0 x^*) >> 0$. Set $y = Ax^*$ and note that $y >> 0$ and $Ay - \lambda_0 y >> 0$. We can assume WLOG that $y'1 = 1$ by normalizing. It follows that there is an $\epsilon > 0$ such that $Ay - (\lambda_0 + \epsilon)y >> 0$. Then $\lambda_0 + \epsilon \in \Lambda$, contradicting the definition of λ_0.

ii) - Let λ be another eigenvalue of A. Then $Az = \lambda z$ and since for any $i, 1 \leq i \leq n$

$$|\lambda||z_i| = |(Az)_i| = |\sum_{j=1}^{n} a_{ij} z_j| \leq \sum_{j=1}^{n} a_{ij}|z_j|$$

it follows that $|\lambda||z| \leq A|z|$, where $|z|$ is the vector of magnitudes of the z's. Since $z \neq 0, |z| > 0$ and we can normalize to find that $|\lambda| \in \Lambda$, from which it follows that $\lambda_0 \geq |\lambda|$.

Suppose that $|\lambda| = \lambda_0$. Let $\delta > 0$ be such that $A_\delta = A - \delta I >> 0$ and consider λ_δ. Since $A_\delta x \geq \eta x$ entails $\eta + \delta \leq \lambda_0$ we have $\lambda_\delta \leq \lambda_0 - \delta$. Taking $x = x^*$ above shows the opposite inequality holds, so $\lambda_\delta = \lambda_0 - \delta$. Observe that $\lambda - \delta$ is an eigenvalue of A_δ to show $|\lambda - \delta| \leq \lambda_0 - \delta = \lambda_\delta$. Therefore, $|\lambda - \delta| \leq |\lambda| - \delta$. Since the opposite inequality holds by the triangle inequality, one has

$$|\lambda - \delta| + \delta = |\lambda|. \tag{1}$$

Here $\delta > 0$ is real while λ may be complex. The set $\{\lambda : |\lambda - \delta| = v > 0\}$ is a circle in the complex plane centered at $(\delta, 0)$ so at whatever value $c = |\lambda|$ at which equality occurs in equation (3), λ must lie simultaneously on the circle of radius c centered at the origin and on the circle of radius $c - \delta$ centered at $(\delta, 0)$. Thus for some θ and $\psi, ce^{i\theta} = (c - \delta)e^{i\psi} + \delta$. Noting that by equation (3) $c \geq \delta$ and dividing both sides by c shows that

$$e^{i\theta} = (1 - \delta/c)e^{i\psi} + \delta/c.$$

The latter convex combination on the rhs of the two points $e^{i\psi}$ and e^{i0} on the unit circle lies strictly inside the circle unless $\psi = 2k\pi$. However, the lhs is on the unit circle in any case. It follows therefore that $e^{i\theta} = 1, e^{i\psi} = 1$, and $\lambda = c = |\lambda| = \lambda_0$.

iii) Suppose we had two linearly independent vectors, x_0 and y, which were eigenvectors for A corresponding to the P-F eigenvalue. Since $A >> 0$ and $x > 0$ entails $Ax >> 0$, it follows that $x_0 >> 0$. Therefore for some $\epsilon > 0, x_0 - \epsilon y >> 0$. We can choose now ϵ large enough so that one or more (but not all since x_0 and y are linearly independent) coordinates of $x_0 - \epsilon y$ are 0. Thus $x_0 - \epsilon y > 0$. But then $A(x_0 - \epsilon y) >> 0$ on the one hand and on the other $A(x_0 - \epsilon y) = \lambda_0(x_0 - \epsilon y)$, so that at least one component is 0, a contradiction.

□

See [2] for a proof of the following.

Theorem 2 *If $A > 0$ and $A^m >> 0$ for some integer $m > 0$ then all of the conclusions of Theorem 1 still hold.*

From the theory of operators (see Rudin, Real and Complex, for example) the spectral radius r of any matrix A, defined by

$$\lim_{n\to\infty} \|A^n\|^{1/n} = \lim_{n\to\infty} (\max_{ij} |a_{ij}(n)|)^{1/n} = r$$

exists and is the magnitude of the largest eigenvalue of A. If A is positive it coincides with the P-F eigenvalue λ_0.

Let $\omega, \omega' A = r\omega'$, be the left P-F eigenvector normalized by $\omega' 1 = 1$ and let $\chi, A\chi = r\chi$, be the right eigenvector normalized by $\omega'\chi = 1$. Define

$$V = \chi\omega'.$$

Some easily checked properties of V are

(i) $V^m = V$,

(ii) $AV = VA = rV$, and

(iii) $(\frac{1}{r}A - V)^m = \frac{1}{r^m}A^m - V$.

Theorem 3 *If $A > 0$ and $A^m >> 0$ for some integer $m > 0$ then*

$$\|\frac{1}{r^m}A^m - V\| \to 0$$

as $m \to \infty$.

Proof: Let $B = \frac{1}{r}A - V$. We claim that if $Bz = \lambda z$ then $|\lambda| < r$. To see this, one has $(\frac{1}{r}A - V)z = \lambda z$ and using (ii) above $\frac{1}{r}(A - AV) = \frac{1}{r}A - V$ so $A(I - V)z = (I - V)Az = r\lambda z$. From the last it follows from $(I - V)^m = I - V$ that

$$A(I - V)z = (I - V)Az = r\lambda(I - V)z. \tag{2}$$

Consider $(I - V)z$. Suppose that it is zero. Then $\omega' z \neq 0$ for if it were 0 we'd have $(I - V)z = z \neq 0$ since z is an eigenvector of B. Therefore, if$(I - V)z = 0$ then $\omega' z \neq 0$. It follows in this case that z is proportional to χ. From the original

$$\frac{1}{r}(A - AV)z = (\frac{1}{r}A - V)z = \lambda z$$

it now follows that λ is 0 and the claim is proven. If $(I - V)z$ is not zero then $r\lambda$ is in fact an eigenvalue of A and it follows from our work above that

$$|\lambda r| < r$$

and hence that the spectral radius t of B, is less than one unless $\lambda = 1$. We show, finally, that $\lambda = 1$ is not possible. If $\lambda = 1$ then

$$A^m(I - V)z = (I - V)Az = r^m(I - V)z.$$

Since $A^m >> 0$ and $\chi >> 0$ there is an $\epsilon > 0$ such that $\chi - \epsilon(I - V)z > 0$ ($(I - V)z$ and χ must be linearly independent for if not then a $\chi = (I - V)z = z - b\chi$ which entails $z = d\chi$ and $\lambda = 0$.) but not $>> 0$. By the usual trick, $A^m(\chi-\epsilon(I-V)z) >> 0$ yet $A^m(\chi-\epsilon(I-V)z) = r^m(\chi-\epsilon(I-V)z) > 0$ and not $>> 0$.

Let $\epsilon > 0$ be such that $t+\epsilon < 1$ and N be sufficiently large so that $n \geq N$ entails $\|B^n\|^{1/n} < t+\epsilon$. Then using iii) one has for $n \geq N$ that

$$\|\frac{1}{r^n}A^n - V\| = \|B^n\| < (t+\epsilon)^n \to 0. \qquad \square$$

Corollary 4 *If $A > 0$ and $A^m >> 0$ for some integer $m > 0$ then given $h > 0$ there is an N such that $n > N$ entails*

$$\|A^n - r^n V\| \leq (|\lambda_2| + h)^n,$$

where λ_2 is the eigenvalue of A second largest in magnitude, so that $|\lambda_2| < r$.

Proof: In Theorem 3 it was shown that the spectral radius t of B is less than 1 and in fact, that if λ is an eigenvalue of B then, except for the eigenvalue r of A, it must be by equation (2) that λr is an eigenvalue of A. It was also shown that $\lambda = 1$ was not possible, so it follows that the eigenvalue of B of greatest magnitude t must satisfy $t = |\lambda_2|/r$, where λ_2 is the eigenvalue of A of second largest magnitude. We saw in Theorem 3 that given $\epsilon > 0$ there is an N such that $n > N$ entails

$$\|\frac{1}{r^n}A^n - V\| = \|B^n\| < (t+\epsilon)^n$$

where t is the spectral radius of B, from which the claim now follows. $\square$

2 Applications

2.1 Limiting distributions of finite state discrete time Markov chains

One says that a Markov chain has a limiting distribution p' if for any vector $a > 0$ with $a'1 = 1$ one has $\lim_{n\to\infty} a'P^n = p'$. The distribution p' is stationary if $p'P = p'$.

Corollary 5 *(Ergodic behavior of Markov chains) Let P be the transition matrix of a Markov chain with finite state space. If $P^m >> 0$ for some $m \geq 1$ then the Perron-Frobenius eigenvalue of P is* 1, *there is a unique limiting distribution p, p' is the left P-F eigenvector and*

$$\frac{\|P^n - 1p'\|}{c^n} \to \begin{cases} 0 & \text{for } c > |t| \\ \infty & \text{for } c < |t| \end{cases}$$

where t is the eigenvalue of P with the second largest magnitude $|t| < 1$.

Proof: P is stochastic so $P1 = 1$. From this and the note from problem 2.2 below it follows that $\lambda = 1$ and χ is proportional to 1. Therefore, consulting Corollary 4, the result is obtained. $\square$

From this it follows that $\|a'P^n - \omega'\| \to 0$ (with rates given above if desired) and hence that P has a limiting distribution. The stationarity of ω' also follows from the fact that it is the left P- F eigenvector.

2.2 Finding r.

Prove that if $A > 0$ and $A^m >> 0$ for some m, r is the Perron-Frobenius eigenvalue, and $x >> 0$ with $x'1 = 1$, then

$$\min_i \sum_{j=1}^n a_{ij}x_j/x_i < r < \max_i \sum_{j=1}^n a_{ij}x_j/x_i$$

or $Ax = rx$.

Proof: Note that for any $x >> 0$ with $x'1 = 1$ one can not have

$$(Ax)_i \geq rx_i \tag{3}$$

for every i with strict inequality for some i. If there were such a vector $z, z'1 = 1$, then one would have $A^m(Az - rz) >> 0$, or setting $y = A^m z >> 0$ and choosing $a = (y'1)^{-1}$, and $v = ay$, one would have $Av >> rv$. This would contradict the definition of r so if (3) holds for every i, then equality holds. Therefore, if $x >> 0$ is not the Perron-Frobenius eigenvector of A, then

$$\min_i \sum_{j=1}^n a_{ij}x_j/x_i < r.$$

Next let ω' be the left eigenvector of A corresponding to r, that is, ω is the P-F eigenvector of A'. If

$$(Ax)_i \leq rx_i \tag{4}$$

for all i then there can be no strict inequality, for if there were one would have $Ax < rx$. But then, since $\omega >> 0, r\omega' x < r\omega' x$, an impossibility. Therefore, if $Ax \neq rx$ and $x >> 0$ with $x'1 = 1$, then (4) must be violated somewhere. Thus

$$r < \max_i \sum_{j=1}^n a_{ij}x_j/x_i. \qquad \square$$

Corollary 6 *If* $x >> 0, x'1 = 1, A > 0,$ *and* $A^m >> 0$ *for some* m, *then* $Ax = \lambda x$ *for some* $\lambda > 0$ *entails* $\lambda = r$ *and* $x = x_0$.

2.3 Game theory

Suppose $A > 0$ and for some integer $m \geq 1, A^m >> 0$. Denoting by $\mathcal{P}$ the set of vectors $x > 0$ in R^n such that $x'1 = 1$ prove that

$$\max_{x\in\mathcal{P}} \min_{y\in\mathcal{P}} \frac{y'Ax}{y'x} = r = \min_{y\in\mathcal{P}} \max_{x\in\mathcal{P}} \frac{y'Ax}{y'x}.$$

Proof: Introduce $\mathcal{Q} = \{x \in \mathcal{P} : x >> 0\}$. Since by problem 2.2 one has $\min_i(\sum_{j=1}^n a_{ij}x_j - rx_i) \leq 0$, it follows from $y \in \mathcal{P}$ that $\min_{y\in\mathcal{P}}(y'Ax - ry'x) \leq 0$. Noting that for $x \in \mathcal{Q}$, $\min_{y\in\mathcal{P}}(y'Ax - ry'x) \leq 0$ entails, $\inf_{y\in\mathcal{P}}(\frac{y'Ax}{y'x} - r) \leq 0$ because $x \in \mathcal{Q}$ and $y \in \mathcal{P}$ implies $y'x > 0$, one has

$$\sup_{x\in\mathcal{Q}} \inf_{y\in\mathcal{P}} \frac{y'Ax}{y'x} \leq r. \tag{5}$$

On the other hand, the P-F eigenvector $\chi = x_0 \in \mathcal{Q}$ and $y \in \mathcal{P}$ entails $\frac{y'Ax_0}{y'x_0} = r$ so

$$\sup_{x\in\mathcal{Q}} \inf_{y\in\mathcal{P}} \frac{y'Ax}{y'x} \geq r. \tag{6}$$

It follows from (5) and (6) that $\sup_{x\in\mathcal{Q}} \inf_{y\in\mathcal{P}} \frac{y'Ax}{y'x} = r$. Furthermore, one has at $x_0 = \chi, r = \inf_{y\in\mathcal{P}} \frac{y'Ax_0}{y'x_0}$ so the supremum is a maximum over $\mathcal{P}$ and since the function $\frac{y'Ax_0}{y'x_0}$ is continuous on $\mathcal{P}$, a compact set, the infimum is a minimum and we have

$$r = \max_{x\in\mathcal{P}} \min_{y\in\mathcal{P}} \frac{y'Ax}{y'x}.$$

Now, in considering the right hand equality, fix $y \in \mathcal{Q}$. Since A' is also positive one has by problem 2.2 applied thereto that $\max_i(\sum_{j=1}^n a'_{ij}y_j - ry_i) \geq 0$, so $\max_{x\in\mathcal{P}}(x'A'y - rx'y) \geq 0$. Since $y \in \mathcal{Q}$ and $x \in \mathcal{P}$ entails $x'y > 0$ it follows that $\sup_{x\in\mathcal{P}}(\frac{y'Ax}{y'x} - r) \geq 0$. Therefore

$$\inf_{y\in\mathcal{Q}} \sup_{x\in\mathcal{P}} \frac{y'Ax}{y'x} \geq r. \tag{7}$$

On the other hand, taking $y = \omega$, the left eigenvector corresponding to the P-F eigenvalue r, one has for any $x \in \mathcal{P}$ that $\frac{\omega'Ax}{\omega'x} = r$ so $\sup_{x\in\mathcal{P}} \frac{\omega'Ax}{\omega'x} = r$. Therefore, since $\omega \in \mathcal{Q}$,

$$\inf_{y\in\mathcal{Q}} \sup_{x\in\mathcal{P}} \frac{y'Ax}{y'x} \leq r. \tag{8}$$

By (7) and (8) one has $r = \inf_{y\in\mathcal{Q}} \sup_{x\in\mathcal{P}} \frac{y'Ax}{y'x}$. Observe that here, as in the case of the maximum, the infimum is attained at $\omega \in \mathcal{P}$ so it is a minimum. Also, the function $\frac{\omega'Ax}{\omega'x}$ is a continuous function on a compact set so the supremum is a maximum and one has

$$r = \min_{y\in\mathcal{P}} \max_{x\in\mathcal{P}} \frac{y'Ax}{y'x}.$$

The proof of the claim is complete. $\square$

Observe that one also has shown that for all $y \in \mathcal{P}$

$$\sup_{x\in\mathcal{P}} \frac{\omega'Ax}{\omega'x} \leq \sup_{x\in\mathcal{P}} \frac{y'Ax}{y'x}$$

and that for all $x \in \mathcal{P}$

$$\inf_{y\in\mathcal{P}} \frac{y'A\chi}{y'\chi} \geq \inf_{y\in\mathcal{P}} \frac{y'Ax}{y'x}$$

so that ω is an optimal strategy for the minimizing player and χ an optimal strategy for the maximizing player in the two-person zero-sum game whose kernel is $\frac{y'Ax}{y'x}$. The value of the game is the P-F eigenvalue r.

Start here.

In fact, this reveals another, slicker, proof of the result above. We have

$$\sup_{x\in\mathcal{P}} \frac{\omega'Ax}{\omega'x} = r = \inf_{y\in\mathcal{P}} \frac{y'A\chi}{y'\chi}$$

from which it follows that

$$\inf_{y\in\mathcal{P}}\sup_{x\in\mathcal{P}}\frac{y'Ax}{y'x}\le r\le\sup_{x\in\mathcal{P}}\inf_{y\in\mathcal{P}}\frac{y'Ax}{y'x}.$$

Since the opposite inequality,

$$\sup_{x\in\mathcal{P}}\inf_{y\in\mathcal{P}}\frac{y'Ax}{y'x}\le\inf_{y\in\mathcal{P}}\sup_{x\in\mathcal{P}}\frac{y'Ax}{y'x}$$

always holds, the result (with infs and sups) is proven along with the value being r and the optimal strategies being identified.

2.4 Random restart in global deterministic optimization

In the study [1] by Hu, Shenk, and Spruill on speedup accorded by random restarting in global optimization, the theory of Perron-Frobenius eigenvalues was shown to be useful. The current state of the global search under stochastic restarting can be modeled as the current state in a Markov chain whose transition matrix is Q, a certain submatrix P of which corresponding to states from which progress can be made towards the global minimum, holds the key to showing that the convergence rate of the probability $\pi' P^n 1$ the goal (minimization state for example) has not been encountered by the nth epoch decreases to 0 geometrically quickly. The authors express the P-F eigenvalue of the deleted transition matrix P (for which $P^m >> 0$ for some $m \ge 1$) as a zero of a polynomial called the structure polynomial. The polynomial depends upon the function to be minimized and the algorithm used to search the values and so reflects the salient features, or *structure* with regard to convergence rate. The eigenvalue and the corresponding eigenvectors are also important in describing the speedup of the parallel restart process towards the goal.

Let $\{X_n : n \ge 1\}$ be a Markov process where $X_1, X_2,$ denote the succession of states in some finite set C, with stationary transition matrix Q. Let D denote a collection of states (to be thought of as the goal states - like points at which the function to be minimized attains its global minimum) in C and suppose there is a collection E of recurrent states (to be thought of as the bottoms of basins in a search for minima at which the algorithm will make no further progress - so a restart is made) in $C - D$. Concerning the initial distribution p' on C, it is assumed that for all $n \ge 1$ and $j \in C - D$

$$p'_j = P[X_1 = j] = P[X_{n+1} = j | X_n \in E].$$

Other transitions are deterministic, dictated by the algorithm (such as steepest descent), the current state in C, and the function being minimized. Denote by π' the deleted probability vector p', with the states of D deleted and by P the deleted transition matrix consisting simply of the elements of Q whose row and column indices are in $C - D$. By the Perron-Frobenius theory given above one has

$$\|P^m - \lambda^m \chi\omega'\| \to 0$$

geometrically fast, where λ is the P-F eignvalue, ω is the left P-F eigenvector normalized so $\omega' 1 = 1$ and χ is the right P-F eignevector normalized so that $\chi'\omega = 1$.

Lemma 2 $\lambda \in (0, 1)$.

Proof: This follows from the fact that P is sub-stochastic ($P1 < 1$ since $\max_{j\in D} p'_j > 0$) by

$$\omega' P 1 = \lambda\omega' 1 = \lambda,$$

since

$$\lambda = \sum_{i=1}^{n} \omega_i (P1)_i < \sum_{i=1}^{n} \omega_i = 1. \qquad \square$$

Let R be the random number of epochs required to first return to the renewal state E given a start in $C - D$. Thus

$$R = \min\{k \geq 0 : X_{n+k} \in E | X_{n-1} \in E\}.$$

Denote by $\phi_R(s)$ the probability generating function of $R, \phi_R(s) = E[s^R]$. and let $\theta = 1 - \pi'1$ be the initial probability of being in C, ie., $\theta = P[X_1 \in C]$ for the original chain. The authors use the facts above to prove, among other things, the following alternative characterization of the rate of convergence.

Theorem 7 *Under the conditions above the P-F eigenvalue of the deleted matrix P is $\lambda = \eta^{-1}$, where $\eta > 1$ is the unique solution to*

$$(1 - \theta)\eta\phi_R(\eta) = 1.$$

2.5 Extension to products with application to simulated annealing

In demonstrating the geometric convergence to 0 as $n \to \infty$ of the probability $\pi' P^n 1$ the goal state had not been encountered by the nth epoch, Corollary 4 was instrumental in the case of deterministic algorithms with restarting. Similar arguments are made by the authors in [3] in the analysis of the convergence to 0 for restarted simulated annealing. In this case the transition matrix is varying with time so the corresponding probability is $\pi' \prod_{j=1}^{n} P_j 1$ and the techniques are different. The key result is Lemma 3 which yields a result analogous to that in Corollary 4 and holds when the time dependent transition matrices P_n converge to a sub-stochastic positive matrix P. See [3] for details.

Lemma 3 *If for some $\tau > 1$, $\sum_{n \geq 1} \tau^n \|P_n - P\| < \infty$ and for some $k \geq 1$, P^k has norm $\eta < 1$, then there is a constant $K < \infty$ such that for all n and m*

$$\|P_m P_{m+1} \cdots P_{m+n-1}\| < K\eta^n.$$

Proof: Let $\delta > 0$ and $1/\beta = \tau - \delta > 1$. By our assumptions there is a finite constant A such that $\|D(j)\| < A(\tau - \delta)^{-j}$ for all j, where $D(j) = P_j - P$. Let M be so large that $m \geq M$ entails $\frac{A\beta^m}{\eta} < 1$. Consider

$$\begin{aligned} P_m P_{m+1} \cdots P_{nk+m-1} - P^{nk} &= (P + D(m))(P + D(m+1)) \cdots (P + D(m+nk-1)) - P^{nk} \\ &= P^{nk-1} D(m+nk-1) + \cdots + D(m) P^{nk-1} \\ &\quad + P^{nk-2} D(m+nk-2) D(m+nk-1) \\ &\quad + \cdots + D(m) D(m+1) P^{nk-2} \\ &\quad + \cdots + D(m) D(m+1) \cdots D(m+nk-1). \end{aligned}$$

The j^{th} term has norm no larger than $K'\eta^n \left(\frac{A\beta^m}{h}\right)^j$ so

$$\|P_m P_{m+1} \cdots P_{nk+m-1} - P^{nk}\| \leq K'\eta^n \sum_{i=1}^{nk} \left(\frac{A\beta^m}{\eta}\right)^i \leq B\eta^n$$

for $m \geq M$ and all n. Now suppose n is arbitrary, $n = dk + \Delta$, where $0 \leq \Delta < k$ so that $\|P^{dk+\Delta} - P_m P_{m+1} \cdots P_{dk+\Delta+m-1}\| = \|AP^{\Delta} - H(P + D(dk + m)) \cdots (P + D(dk + \Delta + m - 1))\|$, where $A = P^{dk}$ and $H = (P + D(m))(P + D(m + 1)) \cdots (P + D(m + dk - 1))$. So

$$\begin{aligned}
\|P^{dk+\Delta} - P_m P_{m+1} \cdots P_{dk+\Delta+m-1}\| &= \|AP^{\Delta} - HP^{\Delta} + HP^{\Delta} - H(P + D(dk + m)) \cdots \\
&\qquad (P + D(dk + \Delta + m - 1))\| \\
&\leq \|A - H\|\|P^{\Delta}\| + \|H\|\|P^{\Delta} - (P + D(dk + m)) \\
&\qquad \cdots (P + D(dk + \Delta + m - 1))\| \\
&\leq B\eta^n \|P^{\Delta}\| + B'\eta^n M \leq K\eta^n.
\end{aligned}$$

References

[1] Hu, Andrea; Shonkwiler, R.; Spruill, M. C. Estimating the convergence rate of a restarted search process. *Int. J. Comput. Numer. Anal. Appl.* 1 (2002), no. 4, 353–367.

[2] Karlin, Samuel; Taylor, Howard M. *A first course in stochastic processes.* Second edition. Academic Press, New York-London, 1975.

[3] Mendivil, F.; Shonkwiler, R.; Spruill, M. C. Restarting search algorithms with applications to simulated annealing. *Adv. in Appl. Probab.* 33 (2001), no. 1, 242–259.