**Theorem 3.** Let \( G^{(t)} \) be the gradient of \( \log \det(\mathbf{A}^T \mathbf{A}) \) \( \text{w.r.t.} \) \( \mathbf{A}^{(t)} \) at iteration \( t \). \( \exists \tau_2 > 0 \), such that \( \forall \eta \in (0, \tau_2) \), \( \Pi(A^{(t+1)}) \leq \Pi(A^{(t)}) \), where \( A^{(t+1)} = \mathcal{P}(A^{(t)} + \eta G^{(t)}) \).

Theorem 3 states that when the step size is sufficiently small, the variance of the pairwise angles decreases. A key observation is that after the updates, \( \theta_{ij}^{(t+1)} - \theta_{ij}^{(t)} = c_{ij} \eta + o(\eta) \), where \( c_{ij} \geq 0 \) and is smaller for larger \( \theta_{ij}^{(t)} \), so that the smaller \( \theta_{ij}^{(t)} \) can "catch up" larger \( \theta_{ij}^{(t)} \). We state that intuition formally in the following lemma, and use it to prove Theorem 3:

**Lemma 3.** Given a non-decreasing sequence \( b = (b_i)_{i=1}^n \) and a strictly decreasing function \( g(x) \) which satisfies \( 0 \leq g(b_i) \leq \min\{b_{i+1} - b_i : i = 1, 2, \cdots, n-1, b_{i+1} \neq b_i\} \), we define a sequence \( c = (c_i)_{i=1}^n \) where \( c_i = b_i + g(b_i) \). If \( b_1 < b_n \), then \( \text{var}(c) < \text{var}(b) \), where \( \text{var}(\cdot) \) denotes the variance of a sequence. Furthermore, let \( n' = \max\{j : b_j \neq b_n\} \), we define a sequence \( b' = (b'_i)_{i=1}^n \) where \( b'_i = b_i + g(b_n) + (g(b_{n'}) - g(b_n))1(i \leq n') \) and \( 1(\cdot) \) is the indicator function, then \( \text{var}(c) \leq \text{var}(b') < \text{var}(b) \).

### 1 Proof of Lemma 3

**Proof.** The intuition behind the proof is that we can view the difference between corresponding elements of the sequence \( b \) and sequence \( c \) as "updates", and we can find that the updates lead to smaller elements "catch up" larger elements. Alternatively, we can obtain the new sequence \( c \) through a set of updates: First, we update the whole sequence \( b \) by the update value of the largest elements, then the largest elements have found their correct values. Then we pick up the elements that are smaller than the largest elements, and update those by the update value of the second largest elements minus the previous update, then the second largest elements have found their correct values. In this manner, we can obtain a sequence of sequences, where the first sequence is \( b \), the third sequence is \( b' \), the last sequence is \( c \), and the adjacent sequences only differ by a simpler update: to the left of some element, each element is updated by a same value; and to the right of the element, each value remains. We can prove that such simpler update can guarantee decreasing of the variance under certain conditions, and we can use that to prove \( \text{var}(c) \leq \text{var}(b') < \text{var}(b) \).

The formal proof starts here: First, following the intuition stated above, we construct a sequence of sequences with decreasing variance, in which the variance of the first sequence is \( \text{var}(b) \) and the variance of the last sequence is \( \text{var}(c) \). We sort the unique values in \( b \) in ascending order and denote the resultant sequence as \( d = (d_j)_{j=1}^m \). Let \( l(j) = \max\{i : b_i = d_j\} \), \( u(i) = \{j : d_j = b_i\} \), we construct a sequence of sequences \( h^{(j)} = (h_i^{(j)})_{i=1}^n \) where \( j = 1, 2, \cdots, m+1 \), in the following way:

1. \( h_i^{(1)} = b_i \), where \( i = 1, 2, \cdots, n \)
2. \( h_i^{(j+1)} = h_i^{(j)} \), where \( j = 1, 2, \cdots, m \) and \( l(m+j) < i \leq n \)
3. \( h_i^{(2)} = h_i^{(1)} + g(d_m) \), where \( 1 \leq i \leq l(m) \)
4. \( h_i^{(j+1)} = h_i^{(j)} + g(d_{m-j+1}) - g(d_{m-j+2}) \), where \( j = 2, 3, \cdots, m \) and \( 1 \leq i \leq l(m+j) \).

From the definition of \( h^{(j)} \), we know \( \text{var}(h^{(1)}) = \text{var}(b) \). As \( b_1 < b_n \), we have \( m \geq 2 \). Now we prove that \( \text{var}(h^{(m+1)}) = \text{var}(c) \) and \( \forall j = 1, 2, \cdots, m \), \( \text{var}(h^{(j+1)}) < \text{var}(h^{(j)}) \).

First, we prove \( \text{var}(h^{(m+1)}) = \text{var}(c) \). Actually, we can prove \( h^{(m+1)} = c \):
\[ h_i^{(m+1)} = \sum_{j=1}^{m} (h_i^{(j+1)} - h_i^{(j)}) + h_i^{(1)} \]
\[ = \sum_{j=1}^{m+1-u(i)} (h_i^{(j+1)} - h_i^{(j)}) + \sum_{j=m+2-u(i)}^{m} (h_i^{(j+1)} - h_i^{(j)}) + b_i \]

As \( j \geq m + 2 - u(i) \iff u(i) \leq m + 2 - j \iff d_{m+2-j} \geq d_{u(i)} = b_i \iff l(m + 1 - j) < i \), we know that

\[ h_i^{(j+1)} = \begin{cases} h_i^{(j)}, & \text{when } j \geq m + 2 - u(i) \\ h_i^{(j)} + g(d_{m-j+1}) - g(d_{m-j+2}), & \text{when } 2 \leq j \leq m + 1 - u(i) \\ h_i^{(j)} + g(d_m), & \text{when } j = 1 \end{cases} \]

So we have

\[ h_i^{(m+1)} = \sum_{j=1}^{m+1-u(i)} (h_i^{(j+1)} - h_i^{(j)}) + b_i \]
\[ = g(d_m) + \sum_{j=2}^{m+1-u(i)} (g(d_{m-j+1}) - g(d_{m-j+2})) + b_i \]
\[ = g(d_m) + g(d_{u(i)}) - g(d_m) + b_i \]
\[ = g(b_i) + b_i \]
\[ = c_i \]

So \( \text{var}(h^{(m+1)}) = \text{var}(c) \).

Then we prove that \( \forall j = 1, 2, \cdots, m, \text{var}(h^{(j+1)}) < \text{var}(h^{(j)}) \). First, we need to prove that for any \( j \), \( h_i^{(j)} \) is a non-decreasing sequence in terms of \( i \). In order to prove that, we only need to prove \( \forall j = 2, \cdots, n, h_i^{(j+1)} - h_i^{(j)} < h_{l^{(m-j+1)+1}}^{(j)} - h_{l^{(m-j+1)+1}}^{(j)} \). Then from \( h_i^{(j+1)} \) to \( h_i^{(j)} \), as elements before \( l(m-j+1) \) are updated by the same value, and elements after \( l(m-j+1) + 1 \) are unchanged, so if the \( l(m-j+1)^{th} \) element does not exceed the \( l(m-j+1) + 1^{st} \) element, then the order of the whole sequence remains during the update. The proof is as follows: \( \forall j \geq 2, h_{l^{(m-j+1)+1}}^{(j)} = \sum_{k=1}^{j-1} (h_{l^{(m-j+1)+1}}^{(k+1)} - h_{l^{(m-j+1)+1}}^{(k)}) + h_{l^{(m-j+1)+1}}^{(1)} \). As \( k \leq j - 1 \Rightarrow l(m-k+1) \geq l(m-j+2) = l(m-j+1+1) \geq l(m-j+1) + 1 \), from the definition of the \( h \) we know that

\[ h_{l^{(m-j+1)+1}}^{(k+1)} - h_{l^{(m-j+1)+1}}^{(k)} = \begin{cases} g(d_{m-k+1}) - g(d_{m-k+2}), & \text{when } k \geq 2 \\ g(d_m), & \text{when } k = 1 \end{cases} \]

So we have

\[ h_{l^{(m-j+1)+1}}^{(j)} = \sum_{k=1}^{j-1} (h_{l^{(m-j+1)+1}}^{(k+1)} - h_{l^{(m-j+1)+1}}^{(k)}) + h_{l^{(m-j+1)+1}}^{(1)} \]
\[ = g(d_m) + \sum_{k=2}^{j-1} (g(d_{m-k+1}) - g(d_{m-k+2})) + b_{l(m-j+1)+1} \]
\[ = g(d_{m-j+2}) + b_{l(m-j+1)+1} \]
From the definition of $I(\cdot)$, we have that $l_{i(m-j+1)+1} = d_{m-j+2}$. So $h_l^{(j)} = g(l_{i(m-j+1)+1}) + b_{l_{i(m-j+1)+1}}$. Similarly, $h_l^{(j)} = b_{l_{i(m-j+1)+1}} + g(b_{l_{i(m-j+1)+2}}) = b_{l_{i(m-j+1)+1}} + g(b_{l_{i(m-j+1)}}) - (g(d_{m-j+1}) - g(d_{m-j+2}))$. So

$$h_l^{(j)} - h_{l_{i(m-j+1)+1}} = b_{l_{i(m-j+1)+1}} - b_{l_{i(m-j+1)}} + (g(b_{l_{i(m-j+1)+1}}) - g(b_{l_{i(m-j+1)}})) + (g(d_{m-j+1}) - g(d_{m-j+2}))$$

As the function $g(x)$ is bounded between 0 and $b_i - b_l$, we have $g(b_{l_{i(m-j+1)+1}}) - g(b_{l_{i(m-j+1)}}) > - (b_{l_{i(m-j+1)+1}} - b_{l_{i(m-j+1)}})$. So

$$h_l^{(j)} - h_{l_{i(m-j+1)+1}} > 0 + (g(d_{m-j+1}) - g(d_{m-j+2})) = g(d_{m-j+1}) - g(d_{m-j+2})$$

Note that $h_l^{(j)} - h_{l_{i(m-j+1)+1}}$ is either 0 or $g(d_{m-j+1}) - g(d_{m-j+2})$ which is positive, we have proved $\forall j \geq 2, h_l^{(j+1)} - h_l^{(j)} < h_{l_{i(m-j+1)+1}} - h_{l_{i(m-j+1)}}$. According to former discussion, for a fixed $j$, $h_l^{(j)}$ is a non-decreasing sequence.

We can prove $\text{var}(h_l^{(j+1)}) < \text{var}(h_l^{(j)})$ now:

If $j = 1$, $l(m) = n$, $\forall i = 1, 2, \cdots, n$, $h_i^{(2)} - h_i^{(1)} = g(d_{m})$, so $\text{var}(h_l^{(2)}) = \text{var}(h_l^{(1)})$.

For $j \geq 2$, let $\Delta_l^{(j)} = g(d_{m-j+1}) - g(d_{m-j+2})$, let $l = l(m-j+1)$, we first use the recursive definition of $h$ to express $h_l^{(j+1)}$ by $h_l^{(j)}$:

$$\text{var}(h_l^{(j+1)}) = \frac{1}{n} \sum_{i=1}^{n} (h_l^{(j+1)} - \frac{1}{n} \sum_{i=1}^{n} h_l^{(j+1)})^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} (h_l^{(j)} + 1(i \leq l)\Delta_l^{(j)}) - \frac{1}{n} \sum_{i=1}^{n} h_l^{(j)} - \frac{l}{n} \Delta_l^{(j)}$$

Then following simple algebra to expand the above equation, we have

$$\text{var}(h_l^{(j+1)}) = \text{var}(h_l^{(j)}) + \frac{\Delta_l^{(j)}}{n} \left[ \frac{1}{l} \sum_{i=1}^{l} h_l^{(j)} - \frac{1}{n} \sum_{i=1}^{n} h_l^{(j)} \right]$$

Note that $h_l^{(j)} - h_l^{(j)} = \Delta_l^{(j)}$, we can further obtain

$$\text{var}(h_l^{(j+1)}) = \text{var}(h_l^{(j)}) + \frac{\Delta_l^{(j)}}{n} \left[ \frac{1}{l} \sum_{i=1}^{l} h_l^{(j)} - \frac{1}{n} \sum_{i=1}^{n} h_l^{(j)} \right]$$

Note that for a fixed $j$, $h_l^{(j)}$ is a non-decreasing sequence, we have $\forall i \geq l+1$, $h_l^{(j)} - h_{l+1}^{(j)} + h_l^{(j)} \geq h_{l+1}^{(j)} - h_l^{(j+1)} + h_l^{(j)} = \Delta_l^{(j)}$, and $\frac{1}{i} \sum_{i=1}^{l} h_l^{(j)} \leq h_l^{(j)}$, so $\frac{1}{l} \sum_{i=1}^{l} h_l^{(j)} - \frac{1}{n} \sum_{i=1}^{n} h_l^{(j)} - \frac{1}{l} \sum_{i=l+1}^{n} (\Delta_l^{(j)} - h_l^{(j)} + h_l^{(j)}) \leq 0$, so $\text{var}(h_l^{(j+1)}) \leq \text{var}(h_l^{(j)}) - \frac{\Delta_l^{(j)}}{n} \frac{l}{n} \Delta_l^{(j)} < \text{var}(h_l^{(j)})$.

Putting above results together, since $\text{var}(h_l^{(j+1)}) < \text{var}(h_l^{(j)})$ and $\text{var}(h_l^{(m+1)}) = \text{var}(c)$, we know that $\text{var}(c) < \text{var}(h_l^{(1)}) = \text{var}(b)$. Furthermore, let $n' = \max \{ j : b_j \neq b_n \}$, then $\forall i, h_i^{(2)} = h_i^{(1)} + g(d_{m}) = b_i + g(b_n)$, $h_i^{(3)} = h_i^{(2)} + g(d_{m}) = b_i + g(b_n) + g(b_{n'}) - g(b_n)) \geq 0$, so $\text{var}(c) \leq \text{var}(b') < \text{var}(b)$. The proof completes.
2 Proof of Theorem 3

Proof. The intuition is that when \( \eta \) is sufficiently small, we can make sure the updates of smaller angles are larger than updates of larger ones, then Lemma 3 can be applied here to prove that the variance decreases. To prove Theorem 3, we need the following conclusion: \( \forall (i, j) \in N \cup V, \theta_{ij}^{(t+1)} - \theta_{ij}^{(t)} = c_{ij} \eta + o(\eta), \) where
\[
c_{ij} = \frac{x_{ij}^{(t)}}{\sqrt{1 - (x_{ij}^{(t)})^2}} \left( \frac{1}{||a_i||} + \frac{1}{||a_j||} \right) = \frac{2 \cos(\theta_{ij}^{(t)})}{\sqrt{1 - \cos(\theta_{ij}^{(t)})}} \text{ if } (i, j) \in N \text{ and } c_{ij} = 0 \text{ if } (i, j) \in V.
\]

In order to apply Lemma 3, we sort \( \theta_{ij}^{(t)} \) in non-decreasing order and denote the resultant sequence as \( \theta^{(t)} = (\theta_{k}^{(t)})_{k=1}^{n}, \) then \( \text{var}((\theta_{ij}^{(t)})) = \text{var}(\theta^{(t)}). \) We use the same order to index \( \theta_{ij}^{(t+1)} \) and denote the resultant sequence as \( \theta_{ij}^{(t+1)} = (\theta_{k}^{(t+1)})_{k=1}^{n}, \) then \( \text{var}((\theta_{ij}^{(t+1)})) = \text{var}(\theta^{(t+1)}). \) Let
\[
g(\theta_{ij}^{(t)}) = \frac{2 \cos(\theta_{ij}^{(t)})}{\sqrt{1 - \cos(\theta_{ij}^{(t)})}} \eta \text{ if } \theta_{ij}^{(t)} < \frac{\pi}{2}
\]
and 0 if \( \theta_{ij}^{(t)} = \frac{\pi}{2}, \) then \( g(\theta_{ij}^{(t)}) \) is a strictly decreasing function. Let \( \frac{\partial}{\partial \theta_{ij}^{(t)}} = \theta_{ij}^{(t)} + c \eta = \theta_{ij}^{(t)} + g(\theta_{ij}^{(t)}). \) It is easy to see when \( \eta \) is sufficiently small, \( 0 \leq g(\theta_{ij}^{(t)}) \leq \min(\theta_{k+1}^{(t)} - \theta_{k}^{(t)} : k = 1, 2, \ldots, n-1, \theta_{k+1}^{(t)} \neq \theta_{k}^{(t)}). \) We continue the proof from two complementary cases: (1) \( \theta_{i}^{(t)} < \theta_{n}^{(t)}; \) (2) \( \theta_{i}^{(t)} = \theta_{n}^{(t)}.
\]

If \( \theta_{i}^{(t)} < \theta_{n}^{(t)} \), then according to Lemma 3, we have \( \text{var}(\tilde{\theta}_{i}^{(t)}) < \text{var}(\theta_{i}^{(t)}), \) where \( \tilde{\theta}_{i}^{(t)} = (\theta_{i}^{(t)})_{k=1}^{n}. \) Furthermore, let \( n' = \max \{ j : \theta_{j}^{(t)} \neq \theta_{n}^{(t)} \}, \theta_{k}^{(t)} = \theta_{k}^{(t)} + g(\theta_{n}^{(t)}) + (g(\theta_{n}^{(t)}) - g(\theta_{k}^{(t)}))1(k \leq n'), \) then \( \text{var}(\tilde{\theta}_{i}^{(t)}) \leq \text{var}(\theta_{i}^{(t)}) < \text{var}(\theta_{i}^{(t)}), \) where \( \theta_{i}^{(t)} = (\theta_{k}^{(t)})_{k=1}^{n}.
\]

We expand \( \text{var}(\theta_{i}^{(t)}) \) below:
\[
\text{var}(\theta_{i}^{(t)}) = \frac{1}{n} \sum_{i=1}^{n} \left( \theta_{i}^{(t)} - \frac{1}{n} \sum_{j=1}^{n} \theta_{j}^{(t)} \right)^2
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} \left( \theta_{i}^{(t)} + g(\theta_{n}^{(t)}) - g(\theta_{n}^{(t)})1(i \leq n') - \frac{1}{n} \sum_{j=1}^{n} \theta_{j}^{(t)} - n' \frac{1}{n} (g(\theta_{n}^{(t)}) - g(\theta_{n}^{(t)})) \right)^2
\]
\[
= \text{var}(\theta_{i}^{(t)}) + 2 \left( \frac{1}{n} \sum_{i=1}^{n} \theta_{i}^{(t)} - \frac{1}{n} \sum_{j=1}^{n} \theta_{j}^{(t)} \right) \left( g(\theta_{n}^{(t)}) - g(\theta_{n}^{(t)})1(i \leq n') - \frac{n'}{n} \right) + \frac{1}{n} \sum_{j=1}^{n} (g(\theta_{n}^{(t)}) - g(\theta_{n}^{(t)}))2(1(i \leq n') - \frac{n'}{n})^2
\]

Let \( \lambda = 2 \left( \frac{1}{n} \sum_{i=1}^{n} \theta_{i}^{(t)} - \frac{1}{n} \sum_{j=1}^{n} \theta_{j}^{(t)} \right) (1(i \leq n') - \frac{n'}{n}) . \) Following simple algebra we have \( \lambda = \frac{2n'(n-n')}{n^2} \left( \frac{1}{n} \sum_{i=1}^{n'} \theta_{i}^{(t)} - \frac{1}{n-n'} \sum_{i=n'+1}^{n} \theta_{i}^{(t)} \right) . \) As \( \theta_{k}^{(t)} \) is non-decreasing and \( \theta_{i}^{(t)} \neq \theta_{n}^{(t)} \), we have \( \lambda < 0. \) Let
\[
\mu = \begin{cases} 
\frac{2 \cos(\theta_{n}^{(t)})}{\sqrt{1 - \cos(\theta_{n}^{(t)})}} - \frac{2 \cos(\theta_{n}^{(t)})}{\sqrt{1 - \cos(\theta_{n}^{(t)})}} , & \text{when } \theta_{n}^{(t)} < \frac{\pi}{2} \\
\frac{2 \cos(\theta_{n}^{(t)})}{\sqrt{1 - \cos(\theta_{n}^{(t)})}} , & \text{when } \theta_{n}^{(t)} = \frac{\pi}{2}
\end{cases}
\]

Then \( g(\theta_{n}^{(t)}) - g(\theta_{n}^{(t)}) = \mu \eta, \) and \( \mu > 0. \) Substituting \( \lambda \) and \( \mu \) into \( \text{var}(\theta_{i}^{(t)}) \), we can obtain
\[
\text{var}(\theta_{i}^{(t)}) = \text{var}(\theta_{i}^{(t)}) + \lambda \eta + \frac{1}{n} \sum_{j=1}^{n} (1(i \leq n') - \frac{n'}{n})^2 \mu^2 \eta^2
\]
\[
= \text{var}(\theta_{i}^{(t)}) + \lambda \eta + o(\eta)
\]

Note that \( \lambda < 0 \) and \( \mu > 0. \) So \( \exists \delta_1 \text{ s.t. } \eta < \delta_1 \Rightarrow \text{var}(\theta_{i}^{(t)}) < \text{var}(\theta_{i}^{(t)}) + \frac{\lambda \eta}{2} \). As \( \text{var}(\tilde{\theta}_{i}^{(t)}) < \text{var}(\theta_{i}^{(t)}) \), we can obtain that \( \text{var}(\tilde{\theta}_{i}^{(t)}) < \text{var}(\theta_{i}^{(t)}) + \frac{\lambda \eta}{2} \eta. \) On the other hand, we can bound the difference between
var(\theta^{(t+1)}) and var(\tilde{\theta}^{(t)}) in the following way:

\[
\text{var}(\theta^{(t+1)}) = \frac{1}{n} \sum_{i=1}^{n} (\theta_i^{(t+1)} - \frac{1}{n} \sum_{j=1}^{n} \theta_j^{(t+1)})^2 = \frac{1}{n} \sum_{i=1}^{n} (\tilde{\theta}_i^{(t)} + o(\eta)) - \frac{1}{n} \sum_{j=1}^{n} (\tilde{\theta}_j^{(t)} + o(\eta))^2
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} (\tilde{\theta}_i^{(t)} - \frac{1}{n} \sum_{j=1}^{n} \tilde{\theta}_j^{(t)})^2 + o(\eta)
\]

\[
= \text{var}(\tilde{\theta}^{(t)}) + o(\eta)
\]

So \( \exists \delta > 0 \) s.t. \( \eta < \delta \Rightarrow \text{var}(\theta^{(t+1)}) < \text{var}(\tilde{\theta}^{(t)}) - \frac{\lambda_2^2}{4} \eta \). Let \( \delta = \min\{\delta_1, \delta_2\} \), then \( \eta < \delta \Rightarrow \text{var}(\theta^{(t+1)}) < \text{var}(\tilde{\theta}^{(t)}) + \frac{\lambda_2^2}{4} \eta < \text{var}(\tilde{\theta}^{(t)}) \Rightarrow \text{var}(\theta^{(t+1)}) < \text{var}(\tilde{\theta}^{(t)}) \).

For the second case \( \theta_1^{(t)} = \theta_0^{(t)} \), i.e. \( \forall (i_1, j_1), (i_2, j_2) \in N \cup V, \theta_{i_1j_1}^{(t)} = \theta_{i_2j_2}^{(t)} \), we prove that \( \text{var}(\theta^{(t+1)}) = \text{var}(\tilde{\theta}^{(t)}) \). In this case, \( \forall (i_1, j_1), (i_2, j_2) \in N \cup V, ((A(t))^T A(t))_{i_1j_1} = ((A(t))^T A(t))_{i_2j_2} \). Denote \( p_1 = a_{i_1}^T \cdot a_{j_1}^T \) for \( i \neq j \) and \( p_2 = a_{i_1}^T \cdot a_{j_1}^T \) for \( i = j \). As \( A^{(t+1)} = A(t) + \eta A(t)((A(t))^T A(t))^{-1} \), we have \( \text{var}(\theta^{(t+1)}) = (A(t))^T A(t) + 2\eta I + \eta^2 ((A(t))^T A(t))^{-1} \). It is apparent that \( \forall (i_1, j_1), (i_2, j_2) \in N \cup V, ((A(t))^T A(t) + 2\eta I)_{i_1j_1} = ((A(t))^T A(t) + 2\eta I)_{i_2j_2} \). For \( \eta^2 ((A(t))^T A(t))^{-1} \), write it as \( \eta^2 ((p_2 - p_1)I_K + p_1 I_K I_K^{-1})^{-1} \), where \( I_K \) is the identity matrix and \( I_K \) is a vector of \( 1 \)s whose length is \( K \). Applying Sherman-Morrison formula, we can obtain that the inverse of \( (A(t))^T A(t) \) is \( \eta^2 ((p_2 - p_1)^{-1} I_K - \frac{p_2 - p_1}{1+\eta p_2} I_K)^{-1} \), we can also see that \( \forall (i_1, j_1), (i_2, j_2) \in N \cup V, ((A(t))^T A(t))_{i_1j_1} = ((A(t))^T A(t))_{i_2j_2} \), so \( ((A(t+1))^T A(t+1))_{i_1j_1} = ((A(t+1))^T A(t+1))_{i_2j_2} \), so var(\theta^{(t+1)}) = 0 = var(\tilde{\theta}^{(t)}) \).

Putting these two cases together, we conclude that \( \exists \tau_2 > 0 \), such that \( \forall \eta \in (0, \tau_2), \Pi(A(t+1)) \leq \Pi(A(t)) \). The proof completes. \( \square \)