



HAL
open science

Harder, Better, Faster, Stronger Convergence Rates for Least-Squares Regression

Aymeric Dieuleveut, Nicolas Flammarion, Francis Bach

► **To cite this version:**

Aymeric Dieuleveut, Nicolas Flammarion, Francis Bach. Harder, Better, Faster, Stronger Convergence Rates for Least-Squares Regression. 2016. hal-01275431v1

HAL Id: hal-01275431

<https://hal.science/hal-01275431v1>

Preprint submitted on 17 Feb 2016 (v1), last revised 23 Feb 2016 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Regularized constant-step size accelerated stochastic gradient descent

February 17, 2016

1 Convergence of Accelerated Averaged Stochastic Gradient Descent

1.1 General expansion

We study the regularized stochastic accelerated gradient descent recursion defined for $n \geq 1$ by

$$\begin{aligned}\theta_n &= \nu_{n-1} - \gamma f'(\nu_{n-1}) - \gamma \lambda (\nu_n - \theta_0) + \gamma \xi_n \\ \nu_n &= \theta_n + \delta(\theta_n - \theta_{n-1}),\end{aligned}$$

starting from $\theta_0 = \nu_0 \in \mathcal{H}$. We may rewrite it for a quadratic function $f : \theta \mapsto \frac{1}{2} \langle \theta - \theta_*, \Sigma(\theta - \theta_*) \rangle$ for $n \geq 2$ as

$$\theta_n = [I - \gamma \Sigma - \gamma \lambda I] [\theta_{n-1} + \delta(\theta_{n-1} - \theta_{n-2})] + \gamma \xi_n + \gamma \lambda \theta_0 + \gamma \Sigma \theta_*,$$

with $\theta_0 \in \mathcal{H}$ and $\theta_1 = [I - \gamma \Sigma - \gamma \lambda I] \theta_0 + \gamma \xi_1 + \gamma \lambda \theta_0 + \gamma \Sigma \theta_*$.

And by centering around the optimum

$$\theta_n - \theta_* = [I - \gamma \Sigma - \gamma \lambda I] [\theta_{n-1} - \theta_* + \delta(\theta_{n-1} - \theta_* - \theta_{n-2} + \theta_*)] + \gamma \xi_n + \lambda \gamma (\theta_0 - \theta_*).$$

With $\delta_n = (1 - 2/n)$ we have for $\eta_n = n(\theta_n - \theta_*)$.

$$\eta_n = [I - \gamma \Sigma - \gamma \lambda I] [2\eta_{n-1} - \eta_{n-2}] + \gamma n \xi_n + \lambda n \gamma (\theta_0 - \theta_*).$$

Thus this is a second order iterative system which is standard to cast in a linear form

$$\Theta_n = F \Theta_{n-1} + \gamma n \Xi_n + \gamma \lambda n \Theta_\lambda, \quad (1)$$

with $T = I - \gamma \Sigma - \gamma \lambda I$, $F = \begin{pmatrix} 2T & -T \\ I & 0 \end{pmatrix}$, $\Theta_n = \begin{pmatrix} n(\theta_n - \theta_*) \\ (n-1)(\theta_{n-1} - \theta_*) \end{pmatrix}$, $\Theta_0 = \begin{pmatrix} \theta_0 - \theta_* \\ 0 \end{pmatrix}$, $\Xi_n = \begin{pmatrix} \xi_n \\ 0 \end{pmatrix}$ and $\Theta_\lambda = \begin{pmatrix} \theta_0 - \theta_* \\ 0 \end{pmatrix}$.

We are interested in the behavior of the average $\bar{\Theta}_n = \frac{1}{(n+1)^2} \sum_{k=0}^n k \Theta_k$ for which we have the following general convergence result:

FB: do you need $\lambda > 0$?

FB: what is $\tilde{\Theta}$? only for Θ_0 . reintroduced in Lemma.

Lemma 1 For all $\lambda \in \mathbb{R}_+$ and γ such that $\gamma(\Sigma + \lambda I) \preceq I$ and any matrix C the average of the iterate Θ_n defined by Eq. (1) satisfy for $P_k \stackrel{(def)}{=} C^{1/2}(I - F^k)(I - F)^{-1}$, with $\tilde{\Theta}_0 = \Theta_0 - \gamma\lambda(I - F)^{-1}\Theta_\lambda$,

$$\begin{aligned} \mathbb{E}\langle \bar{\Theta}_n, C\bar{\Theta}_n \rangle &\leq 2(\gamma\lambda)^2 \|C^{1/2}(I - F)^{-1}\Theta_\lambda\|^2 + \frac{2}{(n+1)^2} \|P_{n+1}\tilde{\Theta}_0\|^2 \\ &\quad + \frac{\gamma^2}{(n+1)^2} \sum_{j=1}^n \text{tr} P_j V P_j^\top. \end{aligned}$$

We remark this is the sum of three terms:

- a regularization based term, which is the first one,
- a bias term, which is the second one
- a variance term. which is the last one. **ad: correct depending on terminology in the paper (2 or 3 terms, bias and reg based together ?)**

We remark that as we have assumed that Σ is invertible, the matrix $I - F$ can be shown to be invertible for all the considered δ .

The regularization-based term will be studied directly whereas the two others will be studied in two stages. First an heuristic will lead to an asymptotic bound then an exact computation will give a non-asymptotic bound. Then using $C = H = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}$ would give a convergence result on the function value and $C = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$ a result on the iterate. The end of the section is devoted to the proof of this lemma.

Proof The sequence Θ_n satisfies a linear recursion, from which we get, for all $n \geq 1$:

$$\Theta_n = F^n \Theta_0 + \gamma \sum_{k=1}^n F^{n-k} k (\Xi_k + \lambda \Theta_\lambda)$$

We study the averaged sequence : $\bar{\Theta}_n = \frac{1}{n+1} \sum_{k=0}^n \Theta_k$. Using the identity $\sum_{k=0}^{n-1} F^k = (I - F^n)(I - F)^{-1}$, we get

$$\bar{\Theta}_n = \frac{1}{n+1} \sum_{k=0}^n F^k \Theta_0 + \frac{\gamma}{n+1} \sum_{k=1}^n \sum_{j=1}^k F^{k-j} j (\Xi_j + \lambda \Theta_\lambda).$$

With

$$\tilde{\Theta}_0 = \Theta_0 - \gamma\lambda(I - F)^{-1}\Theta_\lambda,$$

and $\sum_{k=1}^n (I - F^k) = \sum_{k=0}^n (I - F^k) = [n + 1 - (I - F^{n+1})(I - F)^{-1}]$

$$\begin{aligned}
(n+1)^2 \bar{\Theta}_n &= (I - F^{n+1})(I - F)^{-1} \tilde{\Theta}_0 + \frac{\gamma}{n+1} \sum_{k=1}^n \sum_{j=1}^k F^{k-j} j (\Xi_j + \lambda \Theta_\lambda) \\
&= (I - F^{n+1})(I - F)^{-1} \tilde{\Theta}_0 + \gamma \sum_{j=1}^n \left(\sum_{k=j}^n F^{k-j} \right) j (\Xi_j + \lambda \Theta_\lambda) \quad \text{inverting sums} \\
&= (I - F^{n+1})(I - F)^{-1} \tilde{\Theta}_0 + \gamma \sum_{j=1}^n \left(\sum_{k=0}^{n-j} F^k \right) j (\Xi_j + \lambda \Theta_\lambda) \\
&= (I - F^{n+1})(I - F)^{-1} \tilde{\Theta}_0 + \gamma \sum_{j=1}^n (I - F^{n+1-j})(I - F)^{-1} j (\Xi_j + \lambda \Theta_\lambda) \\
&= (I - F^{n+1})(I - F)^{-1} \tilde{\Theta}_0 + \gamma \sum_{j=1}^n (I - F^j)(I - F)^{-1} (n+1-j) (\Xi_j + \lambda \Theta_\lambda).
\end{aligned}$$

And for the bias term:

$$\begin{aligned}
\sum_{j=1}^n (I - F^j)(I - F)^{-1} (n+1-j) \Theta_\lambda &= \sum_{j=1}^n (n+1-j)(I - F)^{-1} \Theta_\lambda - (n+1) \sum_{j=1}^n F^j (I - F)^{-1} \Theta_\lambda + \sum_{j=1}^n j F^j (I - F)^{-1} \Theta_\lambda \\
&= \frac{n(n+1)}{2} - (n+1) F (I - F^n) (I - F)^{-1}
\end{aligned}$$

Using martingale square moment inequalities which amount to consider $\Xi_i, i = 1, \dots, n$ independent, so that the variance of the sum is the sum of variances, and denoting by $V = \mathbb{E}[\Xi_n \otimes \Xi_n]$ we have for any positive semi-definite C ,

$$\begin{aligned}
\mathbb{E}\langle \bar{\Theta}_n, C \bar{\Theta}_n \rangle &= \left\| C^{1/2} \left(\frac{1}{n+1} (I - F^{n+1})(I - F)^{-1} \tilde{\Theta}_0 + \gamma \lambda (I - F)^{-1} \Theta_\lambda \right) \right\|^2 \\
&\quad + \frac{\gamma^2}{(n+1)^2} \sum_{j=1}^n \text{tr} (I - F^j) (I - F)^{-1} V (I - F^\top)^{-1} (I - F^j)^\top C,
\end{aligned}$$

where $C^{1/2}$ denotes a symmetric square root of C . Define $P_k \stackrel{(def)}{=} C^{1/2} (I - F^k) (I - F)^{-1}$, we have:

$$\begin{aligned}
\mathbb{E}\langle \bar{\Theta}_n, C \bar{\Theta}_n \rangle &= \left\| \frac{1}{n+1} P_{n+1} \tilde{\Theta}_0 + \gamma \lambda C^{1/2} (I - F)^{-1} \Theta_\lambda \right\|^2 + \frac{1}{(n+1)^2} \sum_{j=1}^n \text{tr} P_j V P_j^\top \\
&\leq 2 (\gamma \lambda)^2 \|C^{1/2} (I - F)^{-1} \Theta_\lambda\|^2 + 2 \frac{1}{(n+1)^2} \|P_{n+1} \tilde{\Theta}_0\|^2 \\
&\quad + \frac{\gamma^2}{(n+1)^2} \sum_{j=1}^n \text{tr} P_j V P_j^\top.
\end{aligned}$$

Using Minkowski's inequality and inequality $(a+b)^2 \leq 2(a^2 + b^2)$ for any $a, b \in \mathbb{R}$. ■

1.2 Asymptotic expansion

To give the main terms that we expect, we first provide an asymptotic analysis, which shall only be understood as an insight and is not necessary for the proof. Operator F will have only eigenvalues

smaller than 1, thus $\|F^j\|$ will decrease exponentially to 0 as $j \rightarrow \infty$. The asymptotic analysis relies on ignoring all terms in which F^j appears. We thus approximately have :

$$\begin{aligned}
\mathbb{E}\langle \bar{\Theta}_n, C\bar{\Theta}_n \rangle &\leq 2(\gamma\lambda)^2 \|C^{1/2}(I-F)^{-1}\Theta_\lambda\|^2 + 2\left\|C^{1/2}\frac{1}{n+1}(I-F^{n+1})(I-F)^{-1}\tilde{\Theta}_0\right\|^2 \\
&\quad + \frac{\gamma^2}{(n+1)^2} \sum_{j=1}^n \text{tr}(I-F^j)(I-F)^{-1}V(I-F^\top)^{-1}(I-F^j)^\top C \\
&\approx 2(\gamma\lambda)^2 \|C^{1/2}(I-F)^{-1}\Theta_\lambda\|^2 + 2\left\|C^{1/2}\frac{1}{n+1}(I-F)^{-1}\tilde{\Theta}_0\right\|^2 \\
&\quad + \frac{\gamma^2}{(n+1)^2} \sum_{j=1}^n \text{tr}(I-F)^{-1}V(I-F^\top)^{-1}C
\end{aligned}$$

where, as it has been explained \approx stands for an equality up to terms that will decay exponentially. However, these terms have to be studied very carefully, what will be done in the Section 1.3.

Using the matrix inversion lemma we have for $C = \begin{pmatrix} c & 0 \\ 0 & 0 \end{pmatrix}$,

$$\begin{aligned}
I-F &= \begin{pmatrix} (1+\delta)(\gamma\Sigma + \gamma\lambda I) - \delta I & \delta(I - (\gamma\Sigma + \gamma\lambda I)) \\ -I & I \end{pmatrix} \\
(I-F)^{-1} &= \begin{pmatrix} (\gamma\Sigma + \gamma\lambda I)^{-1} & \delta(I - (\gamma\Sigma + \gamma\lambda I)^{-1}) \\ (\gamma\Sigma + \gamma\lambda I)^{-1} & (1+\delta)I - \delta(\gamma\Sigma + \gamma\lambda I)^{-1} \end{pmatrix} \\
C^{1/2}(I-F)^{-1} &= \begin{pmatrix} c^{1/2}(\gamma\Sigma + \gamma\lambda I)^{-1} & \delta c^{1/2}(I - (\gamma\Sigma + \gamma\lambda I)^{-1}) \\ 0 & 0 \end{pmatrix}.
\end{aligned} \tag{2}$$

Regularization based term. This gives for the regularization based term

$$\begin{aligned}
\|C^{1/2}(I-F)^{-1}\Theta_\lambda\|^2 &= \left\| \begin{pmatrix} c^{1/2}(\gamma\Sigma + \gamma\lambda I)^{-1} & \delta c^{1/2}(I - (\gamma\Sigma + \gamma\lambda I)^{-1}) \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \theta_0 - \theta_* \\ 0 \end{pmatrix} \right\|^2 \\
&= \left(\frac{1}{\gamma}\right)^2 \|(c^{1/2}(\Sigma + \lambda I)^{-1}(\theta_0 - \theta_*))\|^2.
\end{aligned}$$

Bias term. For the bias term we have

$$\begin{aligned}
\tilde{\Theta}_0 &= \Theta_0 - \gamma\lambda(I-F)^{-1}\Theta_\lambda \\
&= \begin{pmatrix} \theta_0 - \theta_* \\ \theta_0 - \theta_* \end{pmatrix} - \gamma\lambda \begin{pmatrix} (\gamma\Sigma + \gamma\lambda I)^{-1} & \delta(I - (\gamma\Sigma + \gamma\lambda I)^{-1}) \\ (\gamma\Sigma + \gamma\lambda I)^{-1} & (1+\delta)I - \delta(\gamma\Sigma + \gamma\lambda I)^{-1} \end{pmatrix} \begin{pmatrix} \theta_0 - \theta_* \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} \theta_0 - \theta_* \\ \theta_0 - \theta_* \end{pmatrix} - \gamma\lambda \begin{pmatrix} (\gamma\Sigma + \gamma\lambda I)^{-1}(\theta_0 - \theta_*) \\ (\gamma\Sigma + \gamma\lambda I)^{-1}(\theta_0 - \theta_*) \end{pmatrix} \\
&= \begin{pmatrix} [I - \lambda(\Sigma + \lambda I)^{-1}](\theta_0 - \theta_*) \\ [I - \lambda(\Sigma + \lambda I)^{-1}](\theta_0 - \theta_*) \end{pmatrix}.
\end{aligned}$$

Thus this gives for the dominant term

$$\begin{aligned}
\|C^{1/2}(I-F)^{-1}\tilde{\Theta}_0\|^2 &= \left\| \begin{pmatrix} c^{1/2}(\gamma\Sigma + \gamma\lambda I)^{-1} & \delta c^{1/2}(I - (\gamma\Sigma + \gamma\lambda I)^{-1}) \\ 0 & 0 \end{pmatrix} \tilde{\Theta}_0 \right\|^2 \\
&= \|(c^{1/2}[(1-\delta)(\gamma\Sigma + \gamma\lambda I)^{-1} + \delta I][I - \lambda(\Sigma + \lambda I)^{-1}](\theta_0 - \theta_*))\|^2.
\end{aligned}$$

And if c commutes with Σ we have the bound for $\delta \in [\frac{1-\sqrt{\gamma\lambda}}{1+\sqrt{\gamma\lambda}}, 1]$

$$\begin{aligned} \left\| C^{1/2}(I-F)^{-1}\tilde{\Theta}_0 \right\|^2 &\leq \left(\frac{(1-\delta)}{\gamma\lambda} + \delta \right) \left\| (c^{1/2}[I - \lambda(\Sigma + \lambda I)^{-1}](\theta_0 - \theta_*) \right\|^2 \\ &\leq \left(\frac{2}{\sqrt{\gamma\lambda}} + 1 \right) \left\| (c^{1/2}[I - \lambda(\Sigma + \lambda I)^{-1}](\theta_0 - \theta_*) \right\|^2. \end{aligned}$$

Variance term. And for the variance term with $V = \begin{pmatrix} v & 0 \\ 0 & 0 \end{pmatrix}$, we have $C^{1/2}(I-F)^{-1}V^{1/2} = \begin{pmatrix} c^{1/2}(\gamma\Sigma + \gamma\lambda I)^{-1}v^{1/2} & 0 \\ 0 & 0 \end{pmatrix}$, and

$$\text{tr } C^{1/2}(I-F)^{-1}V(I-F^\top)^{-1}C^{1/2} = \text{tr } c(\gamma\Sigma + \gamma\lambda I)^{-1}v(\gamma\Sigma + \gamma\lambda I)^{-1}.$$

These would give the three dominant terms. However in order to control the remainders we have to compute more carefully.

1.3 Direct computation without the regularization based term

The computation of the regularization based term being exact we derive now direct computation for the remainders. We assume Σ has eigenvalues (s_i) and we decompose vectors in an eigenvector basis of Σ with $\theta_n^i = p_i^\top \theta_n$ and $\xi_n^i = p_i^\top \xi_n$ and we have the reduced equation:

$$\Theta_{n+1}^i = F_i \Theta_n^i + \gamma \Xi_{n+1}^i.$$

with $\Theta_0^i = \tilde{\Theta}_0^i$, $F_i = \begin{pmatrix} (1+\delta)T_i & -\delta T_i \\ 1 & 0 \end{pmatrix}$, with $T_i = 1 - \gamma s_i - \gamma\lambda$.

Computing $\tilde{\Theta}_0^i$. **FB: why tilde?** $\tilde{\Theta}_0^i = \Theta_0^i - \gamma\lambda(I - F_i)^{-1}\Theta_\lambda^i$, with $\Theta_0^i = \begin{pmatrix} \theta_0^i - \theta_*^i \\ \theta_0^i - \theta_*^i \end{pmatrix}$, $\Theta_\lambda^i = \begin{pmatrix} \theta_0^i - \theta_*^i \\ 0 \end{pmatrix}$ and $(I - F_i)^{-1}$ given in Eq. (2). Thus

$$\tilde{\Theta}_0^i = \begin{pmatrix} \theta_0^i - \theta_*^i \\ \theta_0^i - \theta_*^i \end{pmatrix} - \frac{\gamma\lambda}{(\gamma s_i + \gamma\lambda)} \begin{pmatrix} 1 & \delta(1 - (\gamma s_i + \gamma\lambda)) \\ 1 & -(1 + \delta)(\gamma s_i + \gamma\lambda) - \delta \end{pmatrix} \begin{pmatrix} \theta_0^i - \theta_*^i \\ 0 \end{pmatrix} = \begin{pmatrix} (1 - \frac{\lambda}{\lambda + s_i})(\theta_0^i - \theta_*^i) \\ (1 - \frac{\lambda}{\lambda + s_i})(\theta_0^i - \theta_*^i) \end{pmatrix}.$$

Study of spectrum of F_i . **FB: cite Candes**

Depending on δ , F_i may have two distinct complex eigenvalues of same modulus, only one (double) eigenvalue, or two real eigenvalues. We only consider the two former cases, which we detail bellow.

Indeed, the characteristic polynomial

$$\chi_{F_i}(X) \stackrel{def}{=} \det(XI - F_i) = X^2 - (1 + \delta)(1 - \gamma(s_i + \lambda))X + \delta(1 - \gamma(s_i + \lambda))$$

has discriminant $\Delta_i = (1 - \gamma(s_i + \lambda))((1 + \delta)^2(1 - \gamma(s_i + \lambda)) - 4\delta)$ which is non positive as far as $\delta \in [\delta_-; \delta_+]$, with $\delta_- = \frac{1 - \sqrt{\gamma(s_i + \lambda)}}{1 + \sqrt{\gamma(s_i + \lambda)}}$, $\delta_+ = \frac{1 + \sqrt{\gamma(s_i + \lambda)}}{1 - \sqrt{\gamma(s_i + \lambda)}}$.

1.3.1 Two distinct eigenvalues

We first assume that F_i has two distinct complex eigenvalues $r_{\pm} = \frac{(1+\delta)(1-\gamma(s_i+\lambda)) \pm \sqrt{-1}\sqrt{-\Delta_i}}{2}$ which are conjugate. Thus the roots are of the form $\rho_i e^{\pm i\omega_i}$ with $\rho_i = \sqrt{\delta(1-\gamma(s_i+\lambda))}$, $\cos(\omega_i) = \frac{(1+\delta)(1-\gamma(s_i+\lambda))}{2\rho_i}$, $\omega_i \in [-\pi/2; \pi/2]$ and $\sin(\omega_i) = \frac{\sqrt{-\Delta_i}}{2\rho_i}$.

Let $Q_i = \begin{pmatrix} r_i^- & r_i^+ \\ 1 & 1 \end{pmatrix}$ be the transfer matrix into an eigenbasis of F_i , i.e., $F_i = Q_i D_i Q_i^{-1}$ with $D_i = \begin{pmatrix} r_i^- & 0 \\ 0 & r_i^+ \end{pmatrix}$ and $Q_i^{-1} = \frac{1}{r_i^- - r_i^+} \begin{pmatrix} 1 & -r_i^+ \\ -1 & r_i^- \end{pmatrix}$.

Computing $P_{i,k}$. We first compute the matrix $P_{i,k}$: With $C_i^{1/2} = \begin{pmatrix} \sqrt{c_i} & 0 \\ 0 & 0 \end{pmatrix}$, $C_i^{1/2} Q_i = \begin{pmatrix} r_i^- \sqrt{c_i} & r_i^+ \sqrt{c_i} \\ 0 & 0 \end{pmatrix}$ we have

$$C_i^{1/2} Q_i (I - D_i^k) (I - D_i)^{-1} = \sqrt{c_i} \begin{pmatrix} \frac{1-(r_i^-)^k}{1-r_i^-} r_i^- & \frac{1-(r_i^+)^k}{1-r_i^+} r_i^+ \\ 0 & 0 \end{pmatrix},$$

and

$$\begin{aligned} P_{i,k} &= C_i^{1/2} Q_i (I - D_i^k) (I - D_i)^{-1} Q_i^{-1} = \frac{\sqrt{c_i}}{r_i^- - r_i^+} \begin{pmatrix} \frac{1-(r_i^-)^k}{1-r_i^-} r_i^- & -\frac{1-(r_i^+)^k}{1-r_i^+} r_i^+ & \frac{1-(r_i^+)^k}{1-r_i^+} r_i^- r_i^+ & -\frac{1-(r_i^-)^k}{1-r_i^-} r_i^+ r_i^- \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &= \sqrt{c_i} \begin{pmatrix} \frac{1}{(1-r_i^-)(1-r_i^+)} & \frac{-r_i^+ r_i^-}{(1-r_i^-)(1-r_i^+)} \\ 0 & 0 \end{pmatrix} \\ &\quad - \frac{\sqrt{c_i}}{r_i^- - r_i^+} \begin{pmatrix} \frac{(r_i^-)^{k+1}}{1-r_i^-} & -\frac{(r_i^+)^{k+1}}{1-r_i^+} & \frac{(r_i^+)^{k+1}}{1-r_i^+} r_i^- & -\frac{(r_i^-)^{k+1}}{1-r_i^-} r_i^+ \\ 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

We also have $P_{i,k} = C_i^{1/2} Q_i (I - D_i^k) (I - D_i)^{-1} Q_i^{-1} = \sum_{j=0}^{k-1} R_{i,j}$ with

$$\begin{aligned} R_{i,j} &= C_i^{1/2} Q_i D_i^j Q_i^{-1} \\ &= \sqrt{c_i} \begin{pmatrix} (r_i^-)^{j+1} & (r_i^+)^{j+1} \\ 0 & 0 \end{pmatrix} Q_i^{-1} \\ &= \frac{\sqrt{c_i}}{r_i^- - r_i^+} \begin{pmatrix} (r_i^-)^{j+1} - (r_i^+)^{j+1} & -r_i^+ (r_i^-)^{j+1} + r_i^- (r_i^+)^{j+1} \\ 0 & 0 \end{pmatrix} \end{aligned}$$

which will give looser inequality.

Bias term. Thus, for the bias term :

$$\begin{aligned} P_{i,k} \Theta_0^i &= \sqrt{c_i} \theta_0^i \frac{1 - r_i^+ r_i^-}{(1-r_i^-)(1-r_i^+)} - \frac{\sqrt{c_i} \theta_0^i}{r_i^- - r_i^+} \begin{pmatrix} [(r_i^-)^{k+1} \frac{1-r_i^+}{1-r_i^-} - (r_i^+)^{k+1} \frac{1-r_i^-}{1-r_i^+}] \\ 0 \end{pmatrix} \\ &= \frac{\sqrt{c_i} \theta_0^i}{\sqrt{(1-r_i^-)(1-r_i^+)}} \begin{pmatrix} \frac{[(1-r_i^+ r_i^-) - \rho_i^k A_1]}{\sqrt{(1-r_i^-)(1-r_i^+)}} \\ 0 \end{pmatrix}, \end{aligned}$$

where

$$\rho_i^k A_1 = \frac{(r_i^-)^{k+1} (1-r_i^+)^2 - (r_i^+)^{k+1} (1-r_i^-)^2}{r_i^- - r_i^+}.$$

This can be bound with the following lemma

Lemma 2 For all $\rho \in (0, 1)$ and $\omega \in [-\pi/2; \pi/2]$ and $r^\pm = \rho(\cos(\omega) \pm \sqrt{-1} \sin(\omega))$ we have:

$$\left| \frac{1 - r^+ r^- - \rho^k |A_1|}{|1 - r^+|} \right| \leq \min\{1 + \rho + e^{-1} + 4\rho^k, 1 + \rho + 1 + \sqrt{5}\rho^{k+1}\} \leq 6 \quad (3)$$

We note that the exact constant seems empirically to be 2.

This gives for the bias term

$$\begin{aligned} \|P_{i,k}\Theta_0^i\| &= \frac{\sqrt{c_i(\theta_0^i)}}{\sqrt{(1-r_i^-)(1-r_i^+)}} \left[\frac{1}{\sqrt{(1-r_i^-)(1-r_i^+)}} ((1-r_i^+ r_i^-) - \rho_i^k A_1) \right] \\ &\leq \frac{\sqrt{c_i(\theta_0^i)}}{\sqrt{\gamma(s_i + \lambda)}} (2 + e^{-1} + (1 + \sqrt{5})\rho^k) \\ &\leq (3 + 3\rho^k) \frac{\sqrt{c_i(\theta_0^i)}}{\sqrt{\gamma(s_i + \lambda)}} \\ &\leq 6 \frac{\sqrt{c_i(\theta_0^i)}}{\sqrt{\gamma(s_i + \lambda)}} \end{aligned}$$

Since:

$$\begin{aligned} (1 - r_i^-)(1 - r_i^+) &= 1 - 2\Re r_i^+ + |r_i^+|^2 \\ &= 1 - (1 + \delta)(1 - \gamma(s_i + \lambda)) + \delta(1 - \gamma(s_i + \lambda)) \\ &= \gamma(s_i + \lambda). \end{aligned}$$

We can also have a looser bound using $P_{i,k}\Theta_0^i = \sum_{j=0}^{k-1} R_{i,j}\Theta_0^i$

$$\begin{aligned} R_{i,j}\Theta_0^i &= \frac{\sqrt{c_i}\theta_0^i}{r_i^- - r_i^+} ((1 - r_i^+)(r_i^-)^{j+1} - (1 - r_i^-)(r_i^+)^{j+1}) \\ &= \sqrt{c_i}\theta_0^i \left(\frac{(r_i^-)^{j+1} - (r_i^+)^{j+1}}{r_i^- - r_i^+} - \frac{r_i^+(r_i^-)^{j+1} - r_i^-(r_i^+)^{j+1}}{r_i^- - r_i^+} \right) \\ &= \sqrt{c_i}\theta_0^i \left(\frac{\rho_i^{j+1} \sin(\omega_i(j+1))}{\rho_i \sin(\omega_i)} - \frac{\rho_i e^{i\omega_i} \rho_i^{j+1} e^{-i\omega_i(j+1)} - \rho_i e^{-i\omega_i} \rho_i^{j+1} e^{+i\omega_i(j+1)}}{\rho_i e^{-i\omega_i} - \rho_i e^{i\omega_i}} \right) \\ &= \sqrt{c_i}\theta_0^i \left(\frac{\rho_i^{j+1} \sin(\omega_i(j+1))}{\rho_i \sin(\omega_i)} - \rho_i^{j+1} \frac{e^{-i\omega_i j} - e^{+i\omega_i j}}{e^{-i\omega_i} - e^{i\omega_i}} \right) \\ &= \sqrt{c_i}\theta_0^i \left(\frac{\rho_i^j \sin(\omega_i(j+1))}{\sin(\omega_i)} - \rho_i^{j+1} \frac{\sin(\omega_i j)}{\sin(\omega_i)} \right) \\ &\leq cde\sqrt{c_i}\theta_0^i \end{aligned}$$

replace cde by its true value which will also gives for the bias term

$$\|P_{i,k}\Theta_0^i\| \leq cde\sqrt{c_i}\theta_0^i k$$

and the final bound

$$\|P_{i,k}\Theta_0^i\|^2 \leq \min\left\{36 \frac{c_i(\theta_0^i)^2}{\gamma(s_i + \lambda)}, 6ncde \frac{c_i(\theta_0^i)^2}{\sqrt{\gamma(s_i + \lambda)}}, n^2 cde^2 c_i(\theta_0^i)^2\right\}$$

Proof We have

$$\begin{aligned}
\rho_i^k A_1 &= \frac{(r_i^-)^{k+1}(1-r_i^+)^2 - (r_i^+)^{k+1}(1-r_i^-)^2}{r_i^- - r_i^+} \\
&= -\frac{(r_i^-)^{k+1} - (r_i^+)^{k+1} - 2r_i^- r_i^+ ((r_i^-)^k - (r_i^+)^k) + (r_i^- r_i^+)(r_i^-)^{k-1} - (r_i^+)^{k-1}}{\sin \omega_i \rho_i} \\
&= -\frac{\rho_i^{k+1} \sin((k+1)\omega_i) - 2\rho_i^{k+2} \sin(k\omega_i) + \rho_i^{k+3} \sin((k-1)\omega_i)}{\sin \omega_i}.
\end{aligned}$$

FB: add more details and help the reader

$$\begin{aligned}
-A_1 &= \rho_i \frac{\sin((k+1)\omega_i) - 2\rho_i \sin(k\omega_i) + \rho_i^2 \sin((k-1)\omega_i)}{\rho_i \sin \omega_i} \\
&= \frac{\sin((k+1)\omega_i) - 2\rho_i \sin(k\omega_i) + \rho_i^2 \sin((k-1)\omega_i)}{\sin \omega_i} \\
&= \frac{(\cos(\omega) - \rho)(\sin(k\omega) - \rho \sin((k-1)\omega)) + \cos(k\omega) \sin(\omega) - \rho \cos((k-1)\omega) \sin(\omega)}{\sin \omega_i} \\
&= \frac{(\cos(\omega) - \rho)^2 \sin((k-1)\omega) + (\cos(\omega) - \rho) \sin(\omega) \cos((k-1)\omega) + \cos(k\omega) \sin(\omega) - \rho \cos((k-1)\omega) \sin(\omega)}{\sin \omega_i} \\
&= \frac{(\cos(\omega) - \rho)^2 \sin((k-1)\omega)}{\sin \omega_i} + (\cos(\omega) - \rho) \cos((k-1)\omega) + \cos(k\omega) - \rho \cos((k-1)\omega) \\
&= \frac{(\cos(\omega) - \rho)^2 \sin((k-1)\omega)}{\sin \omega_i} + 2(\cos(\omega) - \rho) \cos((k-1)\omega) + \sin(\omega) \sin((k-1)\omega) \\
|A_1| &\leq k+2 \quad \text{but also} \\
|A_1| &\leq (1-\rho)(k-1) + 4 \quad \text{thus} \\
\rho_i^k |A_1|^2 &\leq 2\rho_i^k (1-\rho)^2 (k-1)^2 + 32 \leq 34
\end{aligned}$$

$$\begin{aligned}
\frac{1 - r_i^+ r_i^-}{|1 - r_i^+|} &= \frac{1 - \rho^2}{|1 - r_i^+|} \leq 1 + \rho \quad \text{exact if } \omega = 0 \\
\frac{-\rho_i^k |A_1|}{|1 - r_i^+|} &= \rho^k \frac{\frac{(\cos(\omega) - \rho)^2 \sin((k-1)\omega)}{\sin \omega_i} + 2(\cos(\omega) - \rho) \cos((k-1)\omega) + \sin(\omega) \sin((k-1)\omega)}{\sqrt{(1 - \rho \cos \omega)^2 + \rho^2 \sin^2(\omega)}} \\
|\rho^k \frac{\frac{(\cos(\omega) - \rho)^2 \sin((k-1)\omega)}{\sin \omega_i}}{\sqrt{(1 - \rho \cos \omega)^2 + \rho^2 \sin^2(\omega)}}| &\leq \rho^k \frac{(\cos(\omega) - \rho) \sin((k-1)\omega)}{\sin \omega_i} \\
&\leq \rho^k (1 - \rho)(k-1) + \rho^k \frac{(\cos(\omega) - 1) \sin((k-1)\omega)}{\sin \omega_i} \\
&\leq \rho^k (1 - \rho)k - (1 - \rho)\rho^k + \rho^k \frac{(\cos(\omega) - 1) \sin((k-1)\omega)}{\sin \omega_i} \\
&\leq \rho^k (1 + (1 - \rho))^k - \rho^k - (1 - \rho)\rho^k + \rho^k \frac{\sin^2(\omega/2)}{\sin \omega_i} \\
&\leq 1 - (1 - \rho)\rho^k \\
2|\rho^k \frac{(\cos(\omega) - \rho) \cos((k-1)\omega)}{\sqrt{(1 - \rho \cos \omega)^2 + \rho^2 \sin^2(\omega)}}| &\leq 2\rho^k \\
|\rho^k \frac{+\sin(\omega) \sin((k-1)\omega)}{\sqrt{(1 - \rho \cos \omega)^2 + \rho^2 \sin^2(\omega)}}| &\leq \rho^k
\end{aligned}$$

FB: add more details and help the reader. This applies to the entire proof..

Thus :

$$\left| \frac{1 - r_i^+ r_i^- - \rho_i^k |A_1|}{|1 - r_i^+|} \right| \leq 1 + \rho + 1 + 3\rho^{k+1}$$

We also have

$$\begin{aligned}
|\rho^k \frac{\frac{(\cos(\omega) - \rho)^2 \sin((k-1)\omega)}{\sin \omega_i}}{\sqrt{(1 - \rho \cos \omega)^2 + \rho^2 \sin^2(\omega)}}| &\leq \rho^k \frac{(\cos(\omega) - \rho) \sin((k-1)\omega)}{\sin \omega_i} \\
&\leq \rho^k (1 - \rho)(k-1) + \rho^k \frac{(\cos(\omega) - 1) \sin((k-1)\omega)}{\sin \omega_i} \\
&\leq (1 - \frac{1}{k+1})^{k+1} - (1 - \rho)\rho^k + \rho^k \frac{(\cos(\omega) - 1) \sin((k-1)\omega)}{\sin \omega_i} \\
&\leq e^{-1} - (1 - \rho)\rho^k + \rho^k \frac{\sin^2(\omega/2)}{\sin \omega_i}
\end{aligned}$$

Using that

$$k \sup_{x \in [0;1]} x^k * (1 - x) = k \frac{1}{k+1} (1 - \frac{1}{k+1})^k = (1 - \frac{1}{k+1})^{k+1} = \exp((k+1) \ln((1 - \frac{1}{k+1}))) \leq e^{-1}$$

leading to

$$\left| \frac{1 - r_i^+ r_i^- - \rho_i^k |A_1|}{|1 - r_i^+|} \right| \leq 1 + \rho + e^{-1} + 4\rho^k$$

We can also change $3\rho^k$ into $\sqrt{5}\rho^k$. We have used that $|(\rho - \cos(\omega))| \leq (1 - \rho \cos(\omega))$. ■

Variance term As for the variance term, with $V_i = \begin{pmatrix} v_i & 0 \\ 0 & 0 \end{pmatrix}$, we have $\text{tr } P_{i,k} V_i P_{i,k} = \left\| P_{i,k} \begin{pmatrix} \sqrt{v_i} \\ 0 \end{pmatrix} \right\|^2$.

$$\begin{aligned} \left\| P_{i,k} \begin{pmatrix} \sqrt{v_i} \\ 0 \end{pmatrix} \right\| &= \frac{\sqrt{v_i c_i}}{(1 - r_i^-)(1 - r_i^+)} \left[1 + \frac{(r_i^-)^{k+1}(1 - r_i^+) - (r_i^+)^{k+1}(1 - r_i^-)}{r_i^+ - r_i^-} \right] \\ &= \frac{\sqrt{v_i c_i}}{\gamma(s_i + \lambda)} \left[1 - \rho_i^k B_{1,k} \right] \end{aligned}$$

where

$$\rho_i^k B_{1,k} = -\frac{(r_i^-)^{k+1}(1 - r_i^+) - (r_i^+)^{k+1}(1 - r_i^-)}{r_i^+ - r_i^-}$$

which we can bound using the following lemma

Lemma 3 For all $\rho \in (0, 1)$ and $\omega \in [-\pi/2; \pi/2]$ and $r^\pm = \rho(\cos(\omega) \pm \sqrt{-1} \sin(\omega))$ we have:

$$\left| \rho_i^k B_{1,k} \right| \leq 1.75 \quad (4)$$

Where we note that the exact majoration seems to be 1.3.

We can also have a looser bound using $P_{i,k} \begin{pmatrix} v_i^{1/2} \\ 0 \end{pmatrix} = \sum_{j=0}^{k-1} R_{i,j} \begin{pmatrix} v_i^{1/2} \\ 0 \end{pmatrix}$ and

$$\begin{aligned} R_{i,j} \begin{pmatrix} v_i^{1/2} \\ 0 \end{pmatrix} &= \frac{\sqrt{c_i v_i}}{r_i^- - r_i^+} ((r_i^-)^{j+1} - (r_i^+)^{j+1}) \\ &= \frac{\sqrt{c_i v_i} \rho_i^{j+1} \sin(\omega_i(j+1))}{\rho_i \sin(\omega_i)} \\ &\leq \sqrt{c_i v_i} (j+1), \end{aligned}$$

and $\left\| P_{i,k} \begin{pmatrix} v_i^{1/2} \\ 0 \end{pmatrix} \right\| \leq \frac{\sqrt{c_i v_i} (k+1)k}{2}$,

And this gives for the Variance term

$$\begin{aligned} \sum_{k=1}^n \text{tr } P_{i,k} V_i P_{i,k} &\leq v c_i \sum_{k=1}^n \min \left\{ \frac{[1 - \rho_i^k B_{1,k}]^2}{\gamma^2(s_i + \lambda)^2}, \frac{[1 - \rho_i^k B_{1,k}] k(k+1)}{2\gamma(s_i + \lambda)}, \frac{k^2(k+1)^2}{4} \right\} \\ &\leq v c_i \min \left\{ \frac{8n}{\gamma^2(s_i + \lambda)^2}, \frac{(n+1)^3}{2\gamma(s_i + \lambda)}, \frac{(n+1)^5}{20} \right\}. \end{aligned}$$

Proof We have

$$\begin{aligned}
\rho_i^k B_{1,k} &= -\frac{(r_i^-)^{k+1}(1-r_i^+) - (r_i^+)^{k+1}(1-r_i^-)}{r_i^+ - r_i^-} \\
&= -\frac{2\Im[(r_i^-)^{k+1}(1-r_i^+)]}{\sqrt{-\Delta_i}} \\
&= -\frac{\Im[\rho_i^k e^{-(k+1)i\omega_i}(1-\rho_i \cos(\omega_i) - i\rho_i \sin(\omega_i))]}{\sin \omega_i \rho_i} \\
&= \rho_i^k \frac{\cos((k+1)\omega_i) \sin(\omega_i) \rho_i + \sin((k+1)\omega_i)(1-\rho_i \cos(\omega_i))}{\sin \omega_i \rho_i} \\
&= \rho_i^k \left[\rho_i \cos((k+1)\omega_i) + (1-\rho_i \cos(\omega_i)) \frac{\sin((k+1)\omega_i)}{\sin \omega_i} \right].
\end{aligned}$$

Moreover,

$$\begin{aligned}
\left| \rho_i^k (1-\rho_i \cos(\omega_i)) \frac{\sin((k+1)\omega_i)}{\sin \omega_i} \right| &= \left| \rho_i^k (1-\rho_i + \rho_i(1-\cos(\omega_i))) \frac{\sin((k+1)\omega_i)}{\sin \omega_i} \right| \\
&\leq \rho_i^k \left| (1-\rho_i) \frac{\sin((k+1)\omega_i)}{\sin \omega_i} \right| + \rho_i^k \left| \rho_i (1-\cos(\omega_i)) \frac{\sin((k+1)\omega_i)}{\sin \omega_i} \right| \\
&\leq \rho_i^k \left| (1-\rho_i)(k+1) \right| + \rho_i^k \left| \rho_i \sin^2\left(\frac{\omega}{2}\right) \frac{1}{2 \cos\left(\frac{\omega}{2}\right) \sin\left(\frac{\omega}{2}\right)} \right| \\
&\leq \rho_i^k (1-\rho_i)k + \rho_i^k (1-\rho) + \rho_i^k \left| \rho_i \sin^2\left(\frac{\omega}{2}\right) \frac{1}{2 \cos\left(\frac{\omega}{2}\right) \sin\left(\frac{\omega}{2}\right)} \right| \\
&\leq (1-(1-\rho_i))^k (1+(1-\rho_i))^k - \rho_i^k + \frac{1}{2(k+1)} + \rho_i^k \left| \frac{\rho_i}{2} \tan\left(\frac{\omega}{2}\right) \right| \\
&\leq (1-(1-\rho_i)^2)^k + \frac{1}{4} + \frac{1}{2} \leq 1 + \frac{1}{4} + \frac{1}{2} - \rho_i^k
\end{aligned}$$

Thus

$$\left| \rho_i^k B_{1,k} \right| = \rho_i^k + 1 + \frac{1}{4} + \frac{1}{2} - \rho_i^k \leq 1 + \frac{1}{4} + \frac{1}{2} = 1.75$$

■

Lemma 1 Finally :

$$\mathbb{E}(\bar{\Theta}_n^i, C\bar{\Theta}_n^i) \leq K_2 \frac{s_i(\theta_0^i)^2}{(s_i + \lambda)\gamma(n+1)^2} + K_1 \frac{vs_i n}{\gamma^2(s_i + \lambda)^2(n+1)^2}$$

With $K_1 = 2.75^2$, $K_2 = (3 + 3\rho^k)^2$.

1.3.2 One coalescent eigenvalue

We assume that F_i has one coalescent eigenvalue $r_i = \frac{(1+\delta)(1-\gamma(s_i+\lambda))}{2}$. Then, with $\delta = \frac{1-\sqrt{\gamma(s_i+\lambda)}}{1+\sqrt{\gamma(s_i+\lambda)}}$, $r_i = \frac{(1+\delta)(1-\gamma(s_i+\lambda))}{2} = 1 - \sqrt{\gamma(s_i+\lambda)}$. Then F_i can be trigonalized as $F_i = Q_i D_i Q_i^{-1}$ with

$Q_i = \begin{pmatrix} r_i & 1 \\ 1 & 0 \end{pmatrix}$, $D_i = \begin{pmatrix} r_i & 1 \\ 0 & r_i \end{pmatrix}$ and $Q_i^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & -r_i \end{pmatrix}$. We note that for all $k \geq 0$, then $D_i^k = r_i^{k-1} \begin{pmatrix} r_i & k \\ 0 & r_i \end{pmatrix}$.

Once again, we first compute $P_{i,k} : (I_2 - D_i)^{-1} = \begin{pmatrix} \frac{1}{1-r_i} & \frac{1}{(1-r_i)^2} \\ 0 & \frac{1}{1-r_i} \end{pmatrix}$ and $(I_2 - D_i^k)(I_2 - D_i)^{-1} = \begin{pmatrix} \frac{1-r_i^k}{1-r_i} & \frac{1-r_i^k}{(1-r_i)^2} - \frac{kr_i^{k-1}}{1-r_i} \\ 0 & \frac{1-r_i^k}{1-r_i} \end{pmatrix}$

Thus with $M_i = \begin{pmatrix} \sqrt{s_i}r_i & \sqrt{s_i} \\ 0 & 0 \end{pmatrix}$ we have $M_i(I_2 - D_i^k)(I_2 - D_i)^{-1} = \sqrt{s_i} \begin{pmatrix} \frac{1-r_i^k}{1-r_i}r_i & \frac{1-r_i^k}{(1-r_i)^2} - \frac{kr_i^k}{1-r_i} \\ 0 & 0 \end{pmatrix}$.

And

$$\begin{aligned} P_{i,k} &= M_i(I_2 - D_i^k)(I_2 - D_i)^{-1}Q_i^{-1} = \sqrt{s_i} \begin{pmatrix} \frac{1-r_i^k}{(1-r_i)^2} - \frac{kr_i^k}{1-r_i} & \frac{1-r_i^k}{1-r_i}r_i - \left(\frac{1-r_i^k}{(1-r_i)^2} - \frac{kr_i^k}{1-r_i}\right)r_i \\ 0 & 0 \end{pmatrix} \\ &= \sqrt{s_i} \begin{pmatrix} \frac{1-r_i^k}{(1-r_i)^2} - \frac{kr_i^k}{1-r_i} & \frac{1-r_i^k}{(1-r_i)^2}(r_i)^2 + \frac{kr_i^{k+1}}{1-r_i} \\ 0 & 0 \end{pmatrix} \\ &= \frac{\sqrt{s_i}}{1-r_i} \begin{pmatrix} \frac{1-r_i^k}{1-r_i} - kr_i^k & -\frac{1-r_i^k}{1-r_i}(r_i)^2 + kr_i^{k+1} \\ 0 & 0 \end{pmatrix} \end{aligned}$$

Thus

$$\begin{aligned} M_i(I_2 - D_i^k)(I_2 - D_i)^{-1}Q_i^{-1}\Theta_0^i &= \frac{\sqrt{s_i}}{1-r_i} \begin{pmatrix} \frac{1-r_i^k}{1-r_i} - kr_i^k & -\frac{1-r_i^k}{1-r_i}(r_i)^2 + kr_i^{k+1} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \theta_0^i \\ \theta_0^i \end{pmatrix} \\ &= \theta_0^i \sqrt{s_i} \begin{pmatrix} (1-r_i^k)\frac{1+r_i}{1-r_i} - kr_i^k \\ 0 \end{pmatrix} \end{aligned}$$

And this gives for the bias term

$$\begin{aligned} \|P_{i,k}\Theta_0^i\|^2 &= (\theta_0^i)^2 s_i \left[(1-r_i^k)\frac{1+r_i}{1-r_i} - kr_i^k \right]^2 \\ &= (\theta_0^i)^2 s_i \left[\frac{1+r_i}{1-r_i} - \left(k + \frac{1+r_i}{1-r_i}\right)r_i^k \right]^2 \\ &= (\theta_0^i)^2 s_i \left[\frac{2 - \sqrt{\gamma(s_i + \lambda)}}{\sqrt{\gamma(s_i + \lambda)}} - \left(k + \frac{2 - \sqrt{\gamma(s_i + \lambda)}}{\sqrt{\gamma(s_i + \lambda)}}\right)(1 - \sqrt{\gamma(s_i + \lambda)})^k \right]^2 \\ &= \frac{(\theta_0^i)^2 s_i}{\gamma(s_i + \lambda)} \left[2 - \sqrt{\gamma(s_i + \lambda)} - (k\sqrt{\gamma(s_i + \lambda)} + 2 - \sqrt{\gamma(s_i + \lambda)})(1 - \sqrt{\gamma(s_i + \lambda)})^k \right]^2 \\ &= \frac{(\theta_0^i)^2 s_i}{\gamma(s_i + \lambda)} \left[2 - \sqrt{\gamma(s_i + \lambda)} - (2 + (k-1)\sqrt{\gamma(s_i + \lambda)})(1 - \sqrt{\gamma(s_i + \lambda)})^k \right]^2 \\ &\leq 4 \frac{(\theta_0^i)^2 s_i}{\gamma(s_i + \lambda)} \end{aligned}$$

As

$$0 \leq 2 - \sqrt{\gamma(s_i + \lambda)} - (2 + (k-1)\sqrt{\gamma(s_i + \lambda)})(1 - \sqrt{\gamma(s_i + \lambda)})^k \leq 2$$

as

$$\begin{aligned}
& \sqrt{\gamma(s_i + \lambda)} + (2 + (k-1)\sqrt{\gamma(s_i + \lambda)})(1 - \sqrt{\gamma(s_i + \lambda)})^k = \\
&= \sqrt{\gamma(s_i + \lambda)} + (1 - \sqrt{\gamma(s_i + \lambda)})^k + (1 + (k-1)\sqrt{\gamma(s_i + \lambda)})(1 - \sqrt{\gamma(s_i + \lambda)})^k \\
&\leq \sqrt{\gamma(s_i + \lambda)} + (1 - \sqrt{\gamma(s_i + \lambda)}) + (1 + (k-1)\sqrt{\gamma(s_i + \lambda)})(1 - \sqrt{\gamma(s_i + \lambda)})^{k-1} \\
&\leq 1 + (1 - \gamma(s_i + \lambda))^{k-1} \leq 2
\end{aligned}$$

And for the variance term, with $V = \begin{pmatrix} v & 0 \\ 0 & 0 \end{pmatrix}$,

$$\begin{aligned}
& \text{tr } P_{i,k} V P_{i,k} \\
&= \frac{s_i}{(1-r_i)^2} \begin{pmatrix} \frac{1-r_i^k}{1-r_i} - kr_i^k & -\frac{1-r_i^k}{1-r_i}(r_i)^2 + kr_i^{k+1} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1-r_i^k}{1-r_i} - kr_i^k & -\frac{1-r_i^k}{1-r_i}(r_i)^2 + kr_i^{k+1} \\ 0 & 0 \end{pmatrix}^\top \\
&= \frac{s_i v}{(1-r_i)^2} \left[\frac{1-r_i^k}{1-r_i} - kr_i^k \right]^2 \\
&= \frac{vh_i}{\gamma(s_i + \lambda)} \left[\frac{1-r_i^k}{1-r_i} - kr_i^k \right]^2 \\
&= \frac{vh_i}{\gamma(s_i + \lambda)(1-r_i)^2} \left[1 - r_i^k - (1-r_i)kr_i^k \right]^2 \\
&= \frac{vh_i}{\gamma^2(s_i + \lambda)^2} \left[1 - (1 + k\sqrt{\gamma(s_i + \lambda)})(1 - \sqrt{\gamma(s_i + \lambda)})^k \right]^2
\end{aligned}$$

And

$$\begin{aligned}
\text{var} &= \sum_{k=1}^n \text{tr } P_{i,k} V P_{i,k} \\
&= \frac{vs_i}{\gamma^2(s_i + \lambda)^2} \sum_{k=1}^n \left[1 - (1 + k\sqrt{\gamma(s_i + \lambda)})(1 - \sqrt{\gamma(s_i + \lambda)})^k \right]^2
\end{aligned}$$

with $x = (1-c)$ and $c = \sqrt{\gamma(s_i + \lambda)}$,

$$\begin{aligned}
\text{var} &= \frac{vs_i}{\gamma^2(s_i + \lambda)^2} \sum_{k=1}^n \left[1 - (1 + kc)(1-c)^k \right]^2 \\
&\leq \frac{vs_i}{\gamma^2(s_i + \lambda)^2} \sum_{k=1}^n \left[1 - (1 + kc)(1-c)^k \right]^2
\end{aligned}$$

It is clear that $(1 + kc)(1-c)^k \leq (1+c)^k(1-c)^k \leq (1-c^2)^k \leq 1$. So

$$\text{var} \leq \frac{vns_i}{\gamma^2(s_i + \lambda)^2}$$

Lemma 2 *We thus have the following conclusion for the coalescing case :*

$$\mathbb{E} \langle \bar{\Theta}_n^i, C \bar{\Theta}_n^i \rangle \leq \frac{4s_i(\theta_0^i)^2}{(s_i + \lambda)\gamma(n+1)^2} + \frac{v's_i}{(n+1)(s_i + \lambda)^2}$$

With $V' = \mathbb{E}\left[\begin{pmatrix} \epsilon_{n+1}^i \\ 0 \end{pmatrix} \otimes \begin{pmatrix} \epsilon_{n+1}^i \\ 0 \end{pmatrix}\right] = \frac{1}{\gamma^2}V$

1.4 Conclusion

Combining Lemma 1 and 2, we have

Theorem 1 *Finally, we get,*

$$\mathbb{E}\langle \bar{\Theta}_n, C\bar{\Theta}_n \rangle \leq 2K_2 \frac{\|\Sigma^{1/2}(\Sigma + \lambda I)^{-1/2}\theta_0\|^2}{\gamma(n+1)^2} + 2\lambda\|(\lambda\Sigma)^{1/2}(\Sigma + \lambda I)^{-1}\theta_*\|^2 + K_1 \frac{\text{tr}(\Sigma(\Sigma + \lambda I)^{-2}V)}{(n+1)}$$

With $K_1 = 2.75^2, K_2 = (3 + 3\rho^k)^2$.