We thank all reviewers for the generally positive comments.

First, we would like to reiterate the contribution of this paper briefly. Covariance matrix estimation is a crucial problem in signal processing but becomes challenging when data evolves rapidly or device resources are limited. Compressive covariance sensing (CCS) addresses this issue by estimating covariance matrices from significantly fewer measurements. This paper focuses on the quadratic sampling model in CCS by assuming a sparse covariance structure. We propose a least-squares estimator with positive definiteness and non-convex sparsity penalties. The estimator is computed via a multistage convex relaxation algorithm based on the majorization-minimization framework. The proposed estimator is theoretically proven to achieve the optimal statistical convergence rate.

In the following, we provide a detailed reply to comments from the reviewers.

Q1: In model (1), $\mathbf{x}_t$ is not observed. (from Reviewers #1) **A1:** Regarding the observation $\mathbf{x}_t$ in model (1), we will change the statement to be $\mathbf{x}_t$ represents the input data.

Q2: In model (1), matrix $\mathbf{S}$ is not defined. (from Reviewers #1) **A2:** While the matrix $\mathbf{S}$ was defined in the introduction, we explicitly restate its definition in model (1) for clarity.

Q3: In (3), the matrix $\boldsymbol{\Sigma}$ is not defined. (from Reviewers #1) **A3:** The matrix $\boldsymbol{\Sigma}$ in (3) is defined as the estimated matrix; we will clarify this in the camera-ready version.

Q4: In definition (7), the norms $\|\cdot\|_0$ and $\|\cdot\|_\mathrm{F}$ are not defined. (from Reviewers #1) **A4:** The norms $\|\cdot\|_0$ and $\|\cdot\|_\mathrm{F}$ are defined as the number of non-zero elements and the Frobenius norm, respectively. The Frobenius norm for a matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$ is given by $\|\mathbf{X}\|_\mathrm{F} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}^2}$.

Q5: Clarify that the quantities $(\boldsymbol{a}_i)_j$ in (2) are the components of the vector $\boldsymbol{a}_i$. (from Reviewers #1) **A5:** In the paper, we denote each element of $\boldsymbol{a}_i$ as $(\boldsymbol{a}_i)_j$ and will provide further clarification on this notation in the final version of the paper.

Q6: Are there alternative algorithms available that could be used for minimizing the same estimator? (from Reviewers #2) **A6:** There are many other algorithms that can solve this problem, but for the sake of statistical proof, we chose the majorization-minimization algorithm.

Q7: It is unclear whether the term "ℓ_1-penalty-based estimator" refers to a distinct estimator or if it simply involves substituting the MCP penalty with the ℓ_1 norm. (from Reviewers #2) **A7:** Thanks for this comment. We will explicitly clarify in the camera-ready version that this referred method is the estimator proposed by [29]. (from Reviewers #2)

Q8: Examples of penalty functions can be provided, but the acronym MCP and the hyperparameter b are used without definition in the paper." (from Reviewers #2 and #3) **A8:** Regarding the definition of MCP, we acknowledge the lack of clarity in its presentation and will revise the paper to ensure precision: "The minimax concave penalty (MCP) is defined as follows: $p_\lambda'(t) = \begin{cases} \lambda - \frac{t}{b}, & t \in [0, b\lambda] \\ 0, & t \in [b\lambda, +\infty) \end{cases}$.

Q9: There is an incomplete sentence on page 4. (from Reviewers #2) **A9:** Thanks for pointing out this issue. We revise it as follows: "Additionally, a comparative analysis is presented between our proposed covariance matrix sensing estimator and a conventional estimator using an ℓ_1 penalty."

Q10: Reference [47] should be revised. (from Reviewers #2 and #3) **A10:** We update the reference [47] to standardize its format.

Q11: Sec III is missing. (from Reviewers #3) **A11:** Section III is complete and not missing.

Q12: The range of ε needs to be discussed. If it is very small, the probability is close to zero. (from Reviewers #3) **A12:** Regarding Theorem 9, we confirm that $\varepsilon \lesssim \sqrt{\frac{1}{mn}}$. Additionally, the probability $1 - \exp\left(-\frac{\epsilon^2}{2}\right)$ involves a different letter ϵ, which only appears visually similar; we may consider changing the notation to avoid confusion.

Q13: What is the meaning of the subscript min in Assumption 6? (from Reviewers #3) **A13:** The subscript $"\min"$ is already defined in Assumption 6.

Q14: Some issues related to the experiments, such as selecting an appropriate title and enhancing the analysis. **A14:** We will enhance the analysis by including more than two values for d in Figure 2. Additionally, we will provide a clearer and more descriptive caption for Figure 3 to convey its content better.

Q15: It needs a proper revision to reflect the contents presented in this conference submission. (from Reviewers #3) **A15:** We will revise Section VII to reflect the contents of this conference submission better and align it with the feedback received.

We will address all remaining minor comments in the final version of the paper.