SuSiE.ASH as a unified framework for fine-mapping and TWAS

Alex McCreight[1,2], Yanghyeon Cho[1,2], FunGen-xQTL Consortium (and/or someone from it)[1], William R.P. Denault[1,*], and Gao Wang[1,5,*]

[1]Center for Statistical Genetics, The Gertrude H. Sergievsky Center, Columbia University, New York, NY, USA
[2]Department of Biostatistics, Columbia University, New York, NY, USA
[3]Additional Co-Author Department, University, City, State, USA
[4]Additional Co-Author Department, University, City, State, USA
[5]Department of Neurology, Columbia University, New York, NY, USA
[*]Correspondence: wang.gao@columbia.edu

ABSTRACT

Many Bayesian variable selection methods, such as Sum of Single Effects Regression (SuSiE), are widely used in genetic association studies to identify causal variants influencing molecular phenotypes like gene expression. These methods oftentimes assume a sparse genetic architecture, implying that only a few variants have large effects on the outcome while the remaining have none. However, this assumption can lead to high false discovery rate (FDR) in settings where multiple variants contribute to the phenotype, including oligogenic and infinitesimal effects. We introduce "SuSiE-ASH (RE) ", a novel method that integrates flexible adaptive shrinkage priors into the SuSiE framework. SuSiE-ASH (RE) offers a flexible alternative to SuSiE by modeling both strong sparse-effect variants and a variety of oligogenic and infinitesimal effects. We conducted extensive simulations reflecting various expression quantitative trait loci (eQTL) settings. Our results deomonstrate that SuSiE-ASH (RE) is able to reduce FDR by up to 50% compared to SuSiE, with only a minimal reduction in power.

METHODS

Model

SuSiE-ASH (RE) is based upon the following model:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{X}\boldsymbol{\theta} + \epsilon, \tag{1}$$

where $\mathbf{y}$ is a centered $n \times 1$ vector of phenotype, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_p]$ is a standarized $n \times p$ matrix of genotypes for p genetic variants in a genomic region of interest, with $\mathbf{x}_j$ being the j-th column of $\mathbf{X}$, the p-vectors β and $\boldsymbol{\theta}$ represent strong spare effect and oligogenic and infinitesimal effects, respectively, which are independent of each other, and $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. Here, we construct β by summing multiple single-effect vectors to model the strong sparse component ([1]). We assume that precisely L variants have a non-zero effect on the outcome:

$$\beta \sim \sum_{\ell=1}^{L} \beta^{(\ell)}, \tag{2}$$

$$\beta^{(\ell)} = (b^{\ell} \odot \gamma^{\ell}), \tag{3}$$

$$\gamma^{\ell} \sim \mathrm{Mult}(1, \mathbf{p}), \tag{4}$$

$$b^{\ell} \sim N(0, \sigma^2_{0\ell}), \tag{5}$$

where $\beta^{(\ell)}$ denotes the ℓ-th single-effect vector, $\gamma^{\ell} = (\gamma_1^{\ell}, \ldots, \gamma_p^{\ell})$ is a p-vector indicating the location of the causal SNP in the ℓ-th single effect, with $\mathbf{p} = (p_1, \ldots, p_p)^T$ representing the prior weight that sum to 1, and b^{ℓ} is a scalar representing the causal effect size in the ℓ-th single effect. Note that the single-effect regression (SER) model is a special case of the above-specified model when $L = 1$.

Then we construct $\boldsymbol{\theta}$ using an adaptive shrinkage prior for the scaled coefficients θ_j/σ ([2]; [3]) to model the remaining oligogenic and infinitesimal effects:

$$\frac{\theta_j}{\sigma} \sim \sum_{k=1}^{K} \pi_k N(0, \sigma_k^2), \tag{6}$$

where $\pi = (\pi_1, \ldots \pi_K)$ represent the mixture proportions which are each non-negative and collectively sum to one, and $\sigma_1^2, \ldots, \sigma_K^2$ are a non-negative, increasing, pre-specified grid of component variances such that $0 \leq \sigma_1^2, \ldots, \sigma_K^2 < \infty$ with σ_1^2 set to 0.

Method

SuSiE-ASH (RE) is a novel approach that improves fine-mapping by combining SuSiE ([1]) for sparse variable selection and Mr.ASH ([2]) for adapative shrinkage estimation. The key idea is to iteratively update the strong sparse effect using SuSiE, and then apply Mr.ASH to the residuals to approximate the remaining oligogenic and infinitesimal effects size variance parameters.

Note that both SuSiE [1] and mr.ASH [2] adopt the variational approximation (VA) method [4] to approximate the posterior distribution under their respective models. By assuming a fully factorized variational approximation, they simplify the optimization of the evidence lower bound (ELBO) over joint prior variables, making it tractable. This tractability is achieved by employing the coordinate ascent algorithm [5], which converts the complex joint optimization problem into a series of simpler tasks, such as fitting single-effect regression (SER) models or normal mean (NM) models.

In SuSiE-ASH (RE) , we have also embraced this approach, assuming that the approximation of the joint posterior q for β is factorized as:

$$q_{\beta}(\beta) = \prod_{\ell=1}^{L} q_{\beta^{\ell}}(\beta^{\ell}). \tag{7}$$

Our proposed Algorithm 1 is iteratively optimizing variational approximations $q_{\beta}(\beta)$, while maximizing the following ELBO under SuSiE-ASH (RE) (1):

$$F(\beta; \sigma^2, \boldsymbol{\sigma}_{0b}^2, \boldsymbol{\sigma}_{\boldsymbol{\theta}}^2, \boldsymbol{\pi}) = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\log\det\Lambda + E_q\left[-\frac{1}{2}(\mathbf{y}-\mathbf{X}\beta)^{T}\Lambda(\mathbf{y}-\mathbf{X}\beta)\right] + E_{q_{\beta}}\left[\log\frac{g_{\beta}(\beta)}{q_{\beta}(\beta)}\right], \tag{8}$$

where $\boldsymbol{\sigma}_{0b}^2 = (\sigma_{01}^2, \ldots, \sigma_{0L}^2)$, $\boldsymbol{\sigma}_{\boldsymbol{\theta}}^2 = (\sigma_1^2, \ldots, \sigma_K^2)$, $\Lambda = (\nu^2\mathbf{X}\mathbf{X}^{T} + \sigma^2\mathbf{I})^{-1}$ with ν^2 = Var(, and g_{β} is the prior distribution of β and $\boldsymbol{\theta}$.

For completeness, the following steps illustrate how Algorithm 1 is implemented within an iteration loop:

Updating the strong spare effect β

Let $\bar{\mathbf{y}}_{\beta}^{(i)} = \mathbf{y} - \mathbf{X}\bar{\theta}^{(i)}$ be the residual after removing the updated oligogenic and infinitesimal effects $\bar{\boldsymbol{\theta}}^{(i)}$ at iteration i. Then, we update $q_{\beta^{\ell}}$ for $\ell = 1, \ldots, L$ by fitting the single-effect regression model:

$$\bar{\mathbf{r}}_{\beta,\ell} := E_{q_{\beta}}\left[\bar{\mathbf{y}}_{\beta}^{(i)} - \sum_{\ell'\neq\ell}\mathbf{X}\beta^{\ell'}\right] = \bar{\mathbf{y}}_{\beta}^{(i)} - \sum_{\ell'\neq\ell}\mathbf{X}\bar{\beta}^{\ell'} = \mathbf{X}\beta^{\ell} + \epsilon. \tag{9}$$

By using the preliminary results from [1], we can easily compute the posterior distribution for $\beta^{\ell} = b^{\ell}\gamma^{\ell}$ under the SER model (Eq. 9):

$$\gamma^{\ell} \mid \mathbf{X}, \bar{\mathbf{y}}_{\beta}^{(i)}, \sigma^{2,(i)}, \boldsymbol{\sigma}_{0b}^2 \sim \text{Mult}(1, \boldsymbol{\alpha}^{\ell}), \tag{10}$$

$$b^{\ell} \mid \mathbf{X}, \bar{\mathbf{y}}_{\beta}^{(i)}, \sigma^{2,(i)}, \boldsymbol{\sigma}_{0b}^2, \gamma_j^{\ell} = 1 \sim N(m_{1j}^{\ell}, \sigma_{1j}^{2,\ell}), \tag{11}$$

where $\boldsymbol{\alpha}^{\ell} = (\alpha_1^{\ell}, \ldots, \alpha_p^{\ell})$ is the vector of the posterior inclusion probabilities (PIPs), with

$$\alpha_j^{\ell} = \frac{p_j \exp\left(\frac{1}{2}\log\frac{\sigma_{1j}^{2,\ell}}{\sigma_{0\ell}^2} + \frac{m_{1j}^{2,\ell}}{2\sigma_{1j}^{2,\ell}}\right)}{\sum_{j'=1}^{p} p_{j'} \exp\left(\frac{1}{2}\log\frac{\sigma_{1j'}^{2,\ell}}{\sigma_{0\ell}^2} + \frac{m_{1j'}^{2,\ell}}{2\sigma_{1j'}^{2,\ell}}\right)} \tag{12}$$

and

$$m_{1j}^{\ell} = \frac{\sigma_{1j}^{2,\ell}}{\sigma^{2,(i)}}\left(\mathbf{X}^{T}\bar{\mathbf{r}}_{\beta,\ell}\right) \text{ and } \sigma_{1j}^{2,\ell} = \left[\frac{\mathbf{x}_j^{T}\mathbf{x}_j}{\sigma^{2,(i)}} + \frac{1}{\sigma_0^{2,(i)}}\right]^{-1} \tag{13}$$

are the posterior mean and variance of b^{ℓ} given $\gamma_j = 1$. Here, $\sigma^2_{(i)}$ is the updated residual variance at iteration i. For the brevity, we introduce the following function that returns arguments of the posterior distribution of β^{ℓ}:

$$\mathrm{SER}(\bar{\mathbf{r}}_{\beta,\ell}, \mathbf{X}; \sigma^{2,(i)}, \boldsymbol{\sigma}^2_{0b}) = (\alpha^{\ell}, \mathbf{m}_1^{\ell}, \boldsymbol{\sigma}_1^{2,\ell}), \tag{14}$$

where $\mathbf{m}_1^{\ell} = (m_{11}^{\ell}, \ldots, m_{1p}^{\ell})$ and $\boldsymbol{\sigma}_1^{2,\ell} = (\sigma_{11}^{2,\ell}, \ldots, \sigma_{1p}^{2,\ell})$.

Updating the oligogenic & spare effect θ

We again redefine the updated residuals, denoted as $\bar{\mathbf{y}}_{\boldsymbol{\theta}}^{(i+1)} = \mathbf{y} - \mathbf{X}\bar{\beta}^{(i+1)}$, by removing the updated spare effect $\bar{\beta}^{(i+1)} = (\bar{\beta}_1^{(i+1)}, \ldots, \bar{\beta}_p^{(i+1)})$, with $\bar{\beta}_j^{(i+1)} = \sum_{\ell}^{L} \alpha_j^{(\ell)} m_{1j}^{\ell}$. In a similar manner with updating the sparse effect, updating oligogenic & inifinitesimal effects involves coordinate-wise approach but each q_{θ_j} for $j = 1, \ldots, p$ is updated by computing a posterior distribution under the following normal mean model:

$$\bar{\mathbf{r}}_{\theta_j}^{(i+1)} := E_{q_{\boldsymbol{\theta}}}\left[\bar{\mathbf{y}}_{\boldsymbol{\theta}}^{(i+1)} - \sum_{j' \neq j} \mathbf{x}_{j'}\theta_{j'}\right] = \bar{\mathbf{y}}_{\boldsymbol{\theta}}^{(i+1)} - \sum_{j' \neq j} \mathbf{x}_{j'}\bar{\mu}_{j'} = \mathbf{x}_j\theta_j + \epsilon. \tag{15}$$

Then, by [2], the posterior distribution for $\boldsymbol{\theta}_j$, denoted as q_{θ_j} under the normal mean model is given by:

$$q(\theta_j|\tilde{\theta}_j; \sigma^2, \boldsymbol{\sigma}^2_{\boldsymbol{\theta}}) = \sum_{k=1}^{K} \phi_{1jk} N(\mu_{1jk}, s^2_{1jk}), \tag{16}$$

where

$$\mu_{1jk} = \frac{\sigma^2\sigma_k^2}{\sigma^2 + \sigma^2\sigma_k^2}\tilde{\theta}_j, \tag{17}$$

$$s^2_{1jk} = \frac{(\sigma^2)^2\sigma_k^2}{\sigma^2 + \sigma^2\sigma_k^2}, \tag{18}$$

$$\phi_{1jk} = \frac{\pi_k L_{jk}}{\sum_{k=1}^{K} \pi_k L_{jk}}, \tag{19}$$

with $\tilde{\theta}_j = (\mathbf{x}_j^T\mathbf{x}_j)^{-1}\mathbf{x}_j^T\bar{\mathbf{r}}_{\theta_j}^{(i+1)}$ and $L_{jk} = N(\tilde{\theta}_j; 0, \sigma^2 + \sigma^2\sigma_k^2)$. For future reference, we define the function $\mathrm{NM}^{\mathrm{post}}$, which returns the estimated parameters for the posterior distribution of θ_j under the normal mean model (Equation 15):

$$\mathrm{NM}^{\mathrm{post}}\left(\bar{\mathbf{r}}_{\theta_j}^{(i+1)}, \mathbf{X}; \sigma^2, \boldsymbol{\sigma}^2_{\boldsymbol{\theta}}\right) = (\boldsymbol{\mu}_{1j}, \mathbf{s}^2_{1j}, \boldsymbol{\phi}_{1j}), \tag{20}$$

where $\boldsymbol{\mu}_{1j} = (\mu_{1j1}, \ldots, \mu_{1jK})$, $\mathbf{s}^2_{1j} = (s^2_{1j1}, \ldots, s^2_{1jK})$, and $\boldsymbol{\phi}_{1j} = (\phi_{1j1}, \ldots, \phi_{1jK})$.

Updating the residual variance σ^2

After updating the sparse effects β and the infinitesimal oligogenic effects $\boldsymbol{\theta}$, SuSiE-ASH (RE) suggests two approaches to estimate σ^2: (1) finding the maximizer of the joint ELBO

(Eq. 21); and (2) applying a method-of-moments (MOM) estimator. Note that

$$
\begin{aligned}
F(\boldsymbol{\beta}, \boldsymbol{\theta}; \sigma^2, \boldsymbol{\sigma}_{0b}^2, \boldsymbol{\sigma}_{\boldsymbol{\theta}}^2) = &-\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}||\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}} - \mathbf{X}\bar{\boldsymbol{\theta}}||^2 - \frac{1}{2\sigma^2}\sum_{j=1}^{p}\mathbf{x}_j^T\mathbf{x}_j \mathrm{Var}_{q_{\boldsymbol{\beta}}}(\beta_j) \\
&- \frac{1}{2\sigma^2}\sum_{j=1}^{p}\mathbf{x}_j^T\mathbf{x}_j \mathrm{Var}_{q_{\boldsymbol{\theta}}}(\theta_j) + \sum_{\ell=1}^{L}\sum_{j=1}^{p}\frac{\alpha_j^{\ell}}{2}\left[1 + \log\frac{\sigma_{1j}^{2,\ell}}{\sigma_{0\ell}^2} - \frac{m_{1j}^{2,\ell} + \sigma_{1j}^{2,\ell}}{\sigma_{0\ell}^2}\right], \\
&- \sum_{j=1}^{p}\left[\sum_{k=1}^{K}\phi_{1jk}\log\frac{\phi_{1jk}}{\pi_k} - \frac{1}{2}\sum_{k=1}^{K}\phi_{1jk}\left\{1 + \log\frac{s_{1jk}^2}{\sigma^2\sigma_k^2} - \frac{s_{1jk}^2 + \mu_{1jk}^2}{\sigma^2\sigma_k^2}\right\}\right]
\end{aligned}
\tag{21}
$$

where $\mathrm{Var}_{q_{\boldsymbol{\beta}}}(\beta_j) = \sum_{\ell}^{L}\left[\alpha_j^{\ell}(m_{1j}^{2,\ell}+\sigma_{1j}^{2,\ell}) - (\alpha_j^{\ell}m_{1j}^{\ell})^2\right]$ and $\mathrm{Var}_{q_{\boldsymbol{\theta}}}(\theta_j) = \sum_{k=1}^{K}\left[\phi_{1jk}(\mu_{1jk}^2+s_{1jk}^2) - (\phi_{1jk}\mu_{1j})^2\right]$. Thus, the closed-form expression for the residual variance estimator using the first approach is obtained by taking the partial derivative of $F(\boldsymbol{\beta}, \boldsymbol{\theta}; \sigma^2, \boldsymbol{\sigma}_{0b}^2, \boldsymbol{\sigma}_{\boldsymbol{\theta}}^2)$ with respect to σ^2:

$$
\sigma_{\mathrm{ELBO}}^{2,(i+1)} = \frac{||\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}} - \mathbf{X}\bar{\boldsymbol{\theta}}||^2 + \sum_{j=1}^{p}\mathbf{x}_j^T\mathbf{x}_j\mathrm{Var}_q(\beta_j) + \sum_{j=1}^{p}\mathbf{x}_j^T\mathbf{x}_j\mathrm{Var}_{q_{\boldsymbol{\beta}}}(\theta_j) + \sum_{j=1}^{p}\sum_{k=2}^{K}\phi_{1jk}(\mu_{1jk}^2 + s_{1jk}^2)/\sigma_k^2}{n + \sum_{j=1}^{p}\sum_{k=2}^{K}\phi_{1jk}}.
\tag{22}
$$

Then, stop iterating if $F(\bar{\boldsymbol{\beta}}^{(i+1)}, \bar{\boldsymbol{\theta}}^{(i+1)}; \sigma_{\mathrm{ELBO}}^{2,(i)}, \boldsymbol{\sigma}_{0b}^2, \boldsymbol{\sigma}_{\boldsymbol{\theta}}^2) - F(\bar{\boldsymbol{\beta}}^{(i)}, \bar{\boldsymbol{\theta}}^{(i)}; \sigma_{\mathrm{ELBO}}^{2,(i-1)}, \boldsymbol{\sigma}_{0b}^2, \boldsymbol{\sigma}_{\boldsymbol{\theta}}^2) < 10^{-3}$.

Alternatively, one may employ the MOM procedure to update the residual variance estimator:

$$
\sigma_{\mathrm{MOM}}^{2,(i+1)} = \frac{E_q||\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\theta}||^2}{n} = \frac{||\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}} - \mathbf{X}\bar{\boldsymbol{\theta}}||^2 + \sum_{j=1}^{p}\mathbf{x}_j^T\mathbf{x}_j\mathrm{Var}_q(\beta_j) + \sum_{j=1}^{p}\mathbf{x}_j^T\mathbf{x}_j\mathrm{Var}_{q_{\boldsymbol{\beta}}}(\theta_j)}{n}.
\tag{23}
$$

Since the MOM estimator does not ensure a non-decreasing ELBO, the convergence criterion for iterations is based on the maximum difference between the posterior inclusion probabilities of the previous and current iterations.

Algorithm 1 SuSiE-ASH Algorithm

Require: $\mathbf{y}, \mathbf{X}, L, K, \sigma^2, \boldsymbol{\sigma}^2_{0b}, \boldsymbol{\sigma}^2_{\boldsymbol{\theta}}$

Initial estimates $\bar{\beta}$, $\bar{\boldsymbol{\theta}}$
while Not converged **do**
    for $\ell = 1$ to L **do**
        $\bar{\mathbf{r}}_{\beta,\ell} = \mathbf{y} - \mathbf{X}\bar{\boldsymbol{\theta}} - \mathbf{X}\sum_{\ell' \neq \ell} \bar{\beta}^{\ell'}$
        $(\boldsymbol{\alpha}^{\ell}, \mathbf{m}_1^{\ell}, \boldsymbol{\sigma}_1^{2,\ell}) \leftarrow \mathrm{SER}(\bar{\mathbf{r}}_{\beta,\ell}, \mathbf{X}; \sigma^2, \boldsymbol{\sigma}^2_{0b})$
        $\bar{\beta}^{\ell} \leftarrow (\boldsymbol{\alpha}^{\ell})^{\mathsf{T}} \mathbf{m}_1^{\ell}$
    end for
    for $j = 1$ to p **do**
        $\bar{\mathbf{r}}_{\theta_j} = \mathbf{y} - \mathbf{X}\bar{\beta} - \sum_{j' \neq j} \mathbf{x}_{j'} \bar{\mu}_{j'}$
        $(\boldsymbol{\mu}_{1j}, \mathbf{s}^2_{1j}, \boldsymbol{\phi}_{1j}) \leftarrow \mathrm{NM}^{\mathrm{post}}(\bar{\mathbf{r}}_{\theta_j}, \mathbf{X}; \sigma^2, \boldsymbol{\sigma}^2_{\boldsymbol{\theta}})$
        $\bar{\mu}_j \leftarrow \sum_{k=1}^{K} \phi_{1jk} \mu_{1jk}$
    end for
    Updating the Residual Variance σ^2 using one of the following:
    - **Maximizing ELBO:** Use Eq. 22
    - **Method-of-Moments:** Use Eq. 23
end while
Output: $\boldsymbol{\alpha}^1, \mathbf{m}_1^1, \boldsymbol{\sigma}_1^{2,1}, \ldots, \boldsymbol{\alpha}^L, \mathbf{m}_1^L, \boldsymbol{\sigma}_1^{2,L}, \bar{\mu}_1, \ldots, \bar{\mu}_p, \sigma^2$

References

1. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. Journal of the Royal Statistical Society Series B: Statistical Methodology *82*, 1273–1300.

2. Kim, Y., Wang, W., Carbonetto, P., and Stephens, M. (2024). A flexible empirical Bayes approach to multiple linear regression and connections with penalized regression. Journal of Machine Learning Research *25*, 1–59.

3. Stephens, M. (2017). False discovery rates: a new deal. Biostatistics *18*, 275–294.

4. Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational inference: A review for statisticians. Journal of the American statistical Association *112*, 859–877.

5. Bertsekas, D. (1999). *Nonlinear Programming*. Athena Scientific.