Simulation Results

The purpose of the simulation study is to investigate the efficacy of MUSS under different magnitudes of measurement error variance and model sparsity by some empirical results.

Model Design

We consider each row of the design matrix X to be generated from $\mathcal{N}(\mathbf{0}, I_p)$. Also, the regression model error vector $\boldsymbol{\varepsilon}$ is drawn from $\mathcal{N}(\mathbf{0}, I_n)$, thus $\sigma^2_{true} = 1$. Each element of the additive noise matrix Ξ is independently generated from $\mathcal{N}(0, \tau)$. The nonzero coefficients are set as 2.5 and appear every 10 coefficients. Let s denote the actual number of nonzero coefficients. When $s = 6$ for instance, the locations of nonzero coefficients are [1,11,21,31,41,51].

Initialization

We set $\boldsymbol{\beta}^{(0)} = \mathbf{0}$ and $\theta^{(0)} = 0.5$. Inappropriate initialization of σ^2 are likely to result the mode stuck in local maximum. In the simulation, we include results from both cases where σ^2 is updated or fixed. For fixed σ^2, we set $\sigma^2 = 1$. For updated σ^2, following Moran, Rockova and George [1] and Chipman, George and McCulloch [2], firstly we compute an over estimated $\hat{\sigma}$ from the standard deviation of $\boldsymbol{y}$, then $\sigma^{(0)}$ is chosen as the mode of the inverse chi-square distribution $3\lambda/\chi^2_3$ where λ is determined such that $P(\sigma < \hat{\sigma}) = 0.99$.

Evaluation Criteria

To evaluate the performance of the algorithm under each specific setting, we run 100 replications and calculate average number of selected variables (q), number of true positive (TP), false discovery rate (FDR), false non-discovery rate (FDR), $\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|_2$ (error1) and $\frac{1}{\sqrt{n}}\left\|X(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\|_2$ (error2).

Results Interpretation

In Table 1, we mainly compare the regression results under increasing measurement error variance. It can be observed that MUSS algorithm generally has better performances in terms of both variable selection and prediction error when σ^2 is fixed as its true value or when prior type is Laplacian. When measurement error is reasonably small, even though σ^2 is updated, both algorithms with Laplacian prior or Gaussian prior could achieve relatively accurate selection results. Particularly, when sample size $n = 200$ and measurement error variance $\tau = 0.5$ or 0.9, $\boldsymbol{q}$'s and TP's are very close to $\boldsymbol{s}$, and the estimation errors (Error1) and prediction errors (Error2) are respectively similar. As the variance of noise gets larger, in this simulation, the number of selected variables declines, thus the false non-discovery rate (FNR) increases conspicuously and Error1 and Error2 grow significantly. Moreover, the influences from prior types and estimation of σ become more noticeable, though larger sample size could ameliorate it to some extent.

In Table 2, by setting $s = 3$ or 10, we investigate the influence of model sparsity on MUSS regression. When true model is sparse ($s = 3$), MUSS could accurately detect nonzero signals with negligible false discoveries and non-discoveries, regardless of the type of priors. Specifically, when $n = 200$, both FDR's and FNR's are exactly 0. As true model gets less sparse ($s = 10$), compared to false discovery, the increased false non-discovery has become more challenging for MUSS. As we can observe, in the setting where $n = 100$, prior type is Gaussian and σ is updated, the average FNR gets to 0.724, while the average FDR still maintains 0.209. Nevertheless, similar to our previous observation in Table 1, if the sample size n is large enough, and it can still be a high-dimensional case, MUSS can still provide satisfying variable selection results.

s	Var(Err)	p	n	Methods	q	TP	FDR	FNR	Error1	Error2
				Laplacian ($\sigma^2 = 1$)	5.93 (0.636)	5.6 (0.693)	0.051 (0.104)	0.067 (0.115)	2.470 (0.777)	2.262 (0.512)
				Laplacian (updated σ)	5.39 (0.989)	5.23 (1.009)	0.031 (0.075)	0.128 (0.168)	2.708 (0.912)	2.489 (0.68)
			100	Gaussian ($\sigma^2 = 1$)	5.46 (1.539)	4.52 (1.1)	0.153 (0.169)	0.247 (0.183)	3.444 (0.951)	2.995 (0.658)
				Gaussian (updated σ)	5.2 (2.117)	4.48 (1.735)	0.120 (0.153)	0.253 (0.289)	3.438 (1.397)	3.072 (1.254)
6	**0.5**	500		Laplacian ($\sigma^2 = 1$)	6.12 (0.354)	6 (0)	0.017 (0.049)	0 (0)	1.862 (0.196)	1.816 (0.145)
				Laplacian (updated σ)	6.07 (0.255)	5.99 (0.099)	0.012 (0.04)	0.002 (0.017)	1.849 (0.223)	1.807 (0.164)
			200	Gaussian ($\sigma^2 = 1$)	6 (0.200)	5.98 (0.140)	0.003 (0.020)	0.003 (0.023)	1.811 (0.238)	1.771 (0.175)
				Gaussian (updated σ)	6.97 (1.118)	5.99 (0.099)	0.121 (0.123)	0.002 (0.017)	1.995 (0.262)	1.88 (0.176)
				Laplacian ($\sigma^2 = 1$)	5.31 (1.659)	4.01 (1.237)	0.216 (0.223)	0.332 (0.206)	4.16 (0.979)	3.651 (0.647)
				Laplacian (updated σ)	3.75 (1.539)	3.31 (1.376)	0.100 (0.169)	0.448 (0.229)	4.404 (0.956)	4.063 (0.801)
			100	Gaussian ($\sigma^2 = 1$)	3.92 (2.521)	2.56 (1.134)	0.264 (0.246)	0.573 (0.189)	5.015 (0.780)	4.395 (0.605)
				Gaussian (updated σ)	2.8 (2.209)	2.18 (1.367)	0.150 (0.242)	0.637 (0.228)	5.203 (0.912)	4.718 (0.844)
6	**0.9**	500		Laplacian ($\sigma^2 = 1$)	6.43 (0.852)	5.96 (0.196)	0.06 (0.105)	0.007 (0.033)	2.587 (0.341)	2.499 (0.227)
				Laplacian (updated σ)	6.09 (0.549)	5.92 (0.306)	0.023 (0.063)	0.013 (0.051)	2.54 (0.304)	2.473 (0.219)
			200	Gaussian ($\sigma^2 = 1$)	5.74 (0.783)	5.48 (0.685)	0.041 (0.079)	0.087 (0.114)	2.819 (0.520)	2.705 (0.419)
				Gaussian (updated σ)	6.6 (1.010)	5.87 (0.416)	0.096 (0.115)	0.022 (0.069)	2.713 (0.374)	2.585 (0.263)
				Laplacian ($\sigma^2 = 1$)	4.78 (1.474)	2.36 (1.091)	0.459 (0.277)	0.607 (0.182)	5.337 (0.721)	4.654 (0.463)
				Laplacian (updated σ)	2.94 (1.489)	1.88 (0.952)	0.26 (0.296)	0.687 (0.159)	5.44 (0.571)	4.971 (0.572)
			100	Gaussian ($\sigma^2 = 1$)	3.24 (2.205)	1.6 (1.058)	0.423 (0.328)	0.733 (0.176)	5.711 (0.697)	5.08 (0.674)
				Gaussian (updated σ)	2.12 (1.739)	1.2 (0.917)	0.29 (0.339)	0.8 (0.153)	5.749 (0.508)	5.332 (0.69)
6	**1.5**	500		Laplacian ($\sigma^2 = 1$)	6.52 (1.136)	5.52 (0.640)	0.135 (0.133)	0.08 (0.107)	3.46 (0.401)	3.308 (0.328)
				Laplacian (updated σ)	4.94 (1.287)	4.7 (1.082)	0.039 (0.090)	0.217 (0.180)	3.779 (0.671)	3.659 (0.582)
			200	Gaussian ($\sigma^2 = 1$)	4.52 (1.220)	4.16 (0.967)	0.065 (0.105)	0.307 (0.161)	3.970 (0.508)	3.788 (0.457)
				Gaussian (updated σ)	4.16 (1.222)	3.86 (1.039)	0.061 (0.106)	0.357 (0.173)	4.141 (0.601)	3.977 (0.585)

Table 1: **Variable selection results under different levels of measurement error. The variance of additive measurement error $\tau = [0.5, 0.9, 1.5]$. Each element of design matrix X is i.i.d. generated from standard normal distribution (IID case). Actual number of nonzero coefficients $s = 6$, total number of potential variables $p = 500$, and sample size $n = 100$ or 200.**

s	Var(Err)	p	n	Methods	q	TP	FDR	FNR	Error1	Error2
				Laplacian (known σ)	3.05 (0.218)	3 (0)	0.013 (0.054)	0 (0)	1.349 (0.177)	1.309 (0.114)
				Laplacian (updated σ)	3.02 (0.140)	3 (0)	0.005 (0.035)	0 (0)	1.345 (0.199)	1.304 (0.150)
			100	Gaussian (known σ)	2.56 (0.605)	2.540 (0.639)	0.010 (0.070)	0.153 (0.213)	1.846 (0.803)	1.732 (0.647)
				Gaussian (updated σ)	2.82 (0.477)	2.760 (0.427)	0.017 (0.067)	0.080 (0.142)	1.628 (0.612)	1.540 (0.491)
3	0.5	500		Laplacian ($\sigma^2 = 1$)	3 (0)	3 (0)	0 (0)	0 (0)	1.250 (0.132)	1.238 (0.104)
				Laplacian (updated σ)	3 (0)	3 (0)	0 (0)	0 (0)	1.283 (0.118)	1.261 (0.105)
			200	Gaussian ($\sigma^2 = 1$)	3 (0)	3 (0)	0 (0)	0 (0)	1.266 (0.119)	1.244 (0.106)
				Gaussian (updated σ)	3 (0)	3 (0)	0 (0)	0 (0)	1.267 (0.119)	1.244 (0.106)
				Laplacian ($\sigma^2 = 1$)	8.49 (1.539)	6.09 (1.733)	0.277 (0.184)	0.391 (0.173)	5.734 (1.203)	4.454 (0.625)
				Laplacian (updated σ)	5.32 (2.195)	4.74 (2.057)	0.104 (0.145)	0.526 (0.206)	6.013 (1.073)	5.239 (0.961)
			100	Gaussian ($\sigma^2 = 1$)	6.76 (2.717)	4.5 (1.552)	0.297 (0.22)	0.55 (0.155)	6.473 (1.027)	5.197 (0.701)
				Gaussian (updated σ)	3.58 (2.426)	2.76 (1.914)	0.209 (0.253)	0.724 (0.191)	7.509 (1.299)	6.438 (1.333)
10	0.5	500		Laplacian ($\sigma^2 = 1$)	10.19 (0.578)	9.91 (0.286)	0.025 (0.047)	0.009 (0.029)	2.657 (0.358)	2.54 (0.277)
				Laplacian (updated σ)	9.93 (0.552)	9.83 (0.401)	0.009 (0.029)	0.017 (0.04)	2.701 (0.434)	2.592 (0.351)
			200	Gaussian ($\sigma^2 = 1$)	9.25 (1.052)	9.08 (0.966)	0.017 (0.043)	0.092 (0.097)	3.135 (0.723)	2.932 (0.576)
				Gaussian (updated σ)	11.23 (1.207)	9.91 (0.286)	0.109 (0.085)	0.009 (0.029)	2.836 (0.377)	2.604 (0.264)

Table 2: **Variable selection results under different model sparsity. The number of true nonzero coefficients is 3 or 10. Each element of design matrix X is i.i.d. generated from standard normal distribution (IID case). The variance of additive measurement error $\tau = 0.5$, total number of potential variables $p = 500$, and sample size $n = 100$ or 200.**

References

[1] Gemma E. Moran, Veronika Rockova, and Edward I. George. Variance prior forms for high-dimensional bayesian variable selection, 2018.

[2] Hugh A. Chipman, Edward I. George, and Robert E. McCulloch. BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266 – 298, 2010.