Target Journal: Journal of Applied Statistics (Review Article)

Tentative Title: Bayesian Modeling of Complex Latent Structures in Regression Analysis

Authors: Chase J. Sakitis[1], Daniel B. Rowe[2]

[1]Division of Health Services & Outcomes Research, Children's Mercy, Kansas City, MO, USA
[2]Department of Mathematical and Statistical Sciences, Marquette University, Milwaukee, WI, USA
None of the authors have conflict-of-interest or financial disclosures to declare.

*Correspondence to:
Chase J. Sakitis, Ph.D.
2401 Gillham Rd
Children's Mercy
Kansas City, MO 64108
cjsakitis@cmh.edu

Daniel B. Rowe – daniel.rowe@marquette.edu

Outline

1. Introduction

1.1. Linear Regression

Linear regression is a common statistical model utilized to predict the behavior of a variable based on other variables. This linear regression model can be expressed with no y-intercept ($\beta_0 = 0$) as shown in Eq. 1.1

$$y_k = \beta_1 x_{k1} + \cdots + \beta_p x_{kp} + \varepsilon_i, \quad k = 1, \ldots, n \qquad [1.1]$$

where p is the number of regression coefficients and n is the number of observations. The model expressed in Eq. 1.1 can be compactly written as

$$y = X\beta + \varepsilon, \qquad [1.2]$$

where $\varepsilon \sim N(0, \sigma^2 I_n)$, $y \in \mathbb{R}^{n\times 1}$ is the observed data, $X \in \mathbb{R}^{n\times p}$ is the design matrix, $\beta \in \mathbb{R}^{p\times 1}$ is the regression coefficients, $\varepsilon \in \mathbb{R}^{n\times 1}$ is the residuals (measurement error), and σ^2 is the variance of the residuals. One purpose of utilizing a linear regression model is to estimate the behavior of the regression coefficients β based on the design matrix X (predictor variables) given the observed data y. Estimating β can also provide hidden results yielded from the observed data and given design matrix. These regression coefficients can be estimated via

$$\hat{\beta} = (X'X)^{-1}(X'y). \qquad [1.3]$$

Utilizing Eq. 1.3 for estimating β is a simple and very common approach when the parameters and data are real-valued and the design matrix is observed or known. However, there are numerous applications that have an unknown design matrix, the parameters are complex-valued instead of real-valued or have both these occurrences.

1.2. Unknown Design Matrix and Complex-valued Applications

Complex-valued latent linear regression with unknown design matrices arises naturally in blind system identification and source separation, where observed data are linear transformations of unobserved sources through an unknown operator. In wireless communications, baseband observations are modeled as linear mixtures of transmitted symbols passed through an unknown channel, yielding bilinear models in which both the regression coefficients and the effective design matrix are latent and motivating blind channel estimation and equalization methods[1,2]. Closely related formulations appear in array (signal) processing, where sensor measurements are linear combinations of latent complex source waveforms mixed through an unknown but structured steering matrix parameterized by physical quantities such as directions of arrival, effectively acting as an unknown design matrix[3,4]. These models connect directly to multichannel signal separation frameworks that treat the mixing operator as a target parameter and rely on structural or statistical assumptions to ensure identifiability[5,6].

From a methodological perspective, much of the blind source separation (BSS) literature can be interpreted as developing estimators for unknown design matrices under increasingly general distributional and dependence assumptions. The independent component analysis (ICA) paradigm formalizes identifiability through independence of latent sources and frames estimation as learning an unknown linear operator via contrast optimization[5]. Subsequent work emphasizes equivariance, stability, and asymptotic behavior of adaptive estimators for unknown mixing matrices[7], as well as unifying analyses of multichannel separation methods based on second-order and higher-order statistics[6]. Extensions to temporally dependent latent processes, such as multichannel ARMA models, further illustrate that uncertainty in the design matrix is often coupled with stochastic structure in the latent regressors rather than simple independent and identically distributed assumptions[8]. Information-theoretic criteria, including information maximization, provide a unifying objective for blind separation and blind deconvolution, while robust and nonlinear learning formulations broaden the class of admissible source distributions and improve stability under model mismatch[9-12].

The complex-valued nature of these models introduces additional statistical considerations, as many practical signals are improper or noncircular, so second-order behavior is not fully characterized by covariance alone. In such settings, consistent estimation requires accounting for augmented second-order structure, with direct implications for identifiability and estimator efficiency when the design matrix is unknown[13]. These issues arise prominently in I/Q baseband systems, array processing, and coherent sensing, where the same physical effects that induce uncertainty in the mixing operator, such as phase drift or calibration error, also generate impropriety in the observed data[13].

Finally, closely related latent-design formulations appear in speech and audio processing, where microphone signals are modeled as mixtures of latent sources convolved with unknown room impulse responses. Blind and semi-blind separation methods are therefore evaluated under strong model uncertainty, emphasizing robustness to reverberation and nonstationarity[14], while speech enhancement frameworks treat inference under uncertain mixing and noise models with perceptual constraints[15]. Sparse representation–based post-processing approaches further exploit latent structure to improve perceptual quality after separation without altering the underlying blind mixing model[16,17]. Related perspectives also arise in coherent imaging and radar autofocus, where unknown calibration or phase errors are naturally modeled as latent linear operators acting on complex-valued signals, leading to blind deconvolution and phase retrieval formulations with incomplete observations[18-20].

1.3. Overview

In these real-world cases of not having a known design matrix and complex-valued data, along with β, we will also need to estimate X and account for complex-valued nature of the data in these applications. Here, we will explain how to handle this linear regression model in these scenarios building up from a real-valued linear regression model with an observed X to a complex-valued linear regression model with an unobserved X. In Section 2, we will discuss the standard linear regression model with an observed design matrix. We will then discuss the linear regression model when the design matrix is unobserved. In Section 4, we will introduce complex-valued parameters and how we handle the latent regression model. We will then finish with a discussion in Section 5.

2. Model 1: Real-Valued Parameters with an Observed Design Matrix

In this first model, we will describe the linear regression model with real-valued parameters and an observed design matrix. In this “textbook scenario”, the observed parameters are dependent data y and the design matrix X. The straightforward approach to estimating β given the observed data is to utilize Eq. 1.3.

2.1. Bayesian Approach to Simple Linear Regression

For a Bayesian approach to this model, we will first need to define the data likelihood, prior distributions, and the posterior distribution. From these, we can then derive posterior marginals and posterior conditionals to estimate the parameters of interest from the posterior distribution. Since the observed data y is assumed to follow a normal distribution, the data likelihood is expressed as

$$f(y|X,\beta,\sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right]. \qquad [2.1]$$

For this model, the regression coefficients β and the variance of the residuals σ^2 are the unobserved parameters that need to be estimated. This requires prior distributions to be placed on both parameters. These prior distributions are

$$f\left(\beta|\beta_0,\sigma^2,n_\beta\right) \propto (\sigma^2)^{-\frac{p}{2}} \exp\left[-\frac{n_\beta}{2\sigma^2}(\beta-\beta_0)'(\beta-\beta_0)\right], \qquad [2.2]$$

$$f(\sigma^2|\alpha,\gamma) \propto (\sigma^2)^{-(\alpha+1)} \exp\left[-\frac{\gamma}{\sigma^2}\right]. \qquad [2.3]$$

Eqs. 2.2 and 2.3 exhibit a normal distribution for β and an inverse gamma distribution for σ^2, and are conjugate priors, simplifying the determination of the posterior distribution. Combining the data likelihood and prior distributions, we determine the posterior to be

$$f(\beta,\sigma^2|X,y) \propto (\sigma^2)^{-\left(\frac{n+p+2\alpha}{2}+1\right)} \exp\left[-\frac{h}{2\sigma^2}\right], \qquad [2.4]$$

where $h = (y-X\beta)'(y-X\beta) + n_\beta(\beta-\beta_0)'(\beta-\beta_0) + 2\gamma$.

2.2. Posterior Marginals and Conditionals

For the posterior marginals, we aim to integrate out the other unobserved parameters from the posterior, leaving the parameter we are currently estimating. So, for β, we would integrate out σ^2 from the posterior, leaving only observed data to estimate β and then vice versa for σ^2. This will give us

$$f(\beta|X,y) \propto \left[1+\frac{1}{\nu}(\beta-\hat{\beta})'\left[\frac{\nu(X'X+n_\beta I_p)}{\phi}\right](\beta-\hat{\beta})\right]^{-\frac{\nu+p}{2}}, \qquad [2.5]$$

$$f(\sigma^2|X,y) \propto (\sigma^2)^{-(\alpha+1)} \exp\left[-\frac{\gamma_*}{\sigma^2}\right], \qquad [2.6]$$

where β follows a multivariate location-scale *t* distribution such that $\beta|X,y \sim t\left(\nu,\hat{\beta},\mathrm{T}\right)$, $E(\beta|X,y) = \hat{\beta} = \left(X'X+n_\beta I_p\right)^{-1}\left(X'y+n_\beta\beta_0\right)$, $Var(\beta|X,y) = \frac{\nu}{\nu-2}T$, and $\nu = 2\alpha+n$ and σ^2 follows an inverse gamma distribution such that $\sigma^2\ |X,y \sim IG\left(\alpha_* = \frac{n+p+2\alpha}{2}, \gamma_* = \frac{\phi}{2}\right)$, $E(\sigma^2|X,y) = \frac{\gamma_*}{\alpha_*-1}$, $Var(\sigma^2|X,y) = \frac{\gamma_*^2}{(\alpha_*-1)^2(\alpha_*-2)}$, and $\phi = y'y - \hat{\beta}'\left(X'X+n_\beta I_p\right)\hat{\beta} + n_\beta{\beta_0}'\beta_0 + 2\gamma$. Since X and y in the expected value equations ($E(\beta|X,y)$ and $E(\sigma^2|X,y)$) are known, we can simply calculate the expected value for β and σ^2 without requiring any computational techniques.

For the posterior conditionals, instead of integrating out β and σ^2, we can leave the unobserved parameters in, mathematically determine the posterior conditionals, and utilize a computational technique to estimate β and σ^2. These yields

$$f(\beta|X,\sigma^2,y) \propto \left|\Sigma_\beta\right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\beta-\hat{\beta})'\left(\Sigma_\beta\right)^{-1}(\beta-\hat{\beta})\right], \qquad [2.7]$$

$$f(\sigma^2|\beta, X, y) \propto (\sigma^2)^{-(\alpha_*+1)} \exp\left[-\frac{\gamma_{**}}{\sigma^2}\right], \qquad [2.8]$$

where β follows a multivariate normal distribution such that $\beta|X, \sigma^2, y \sim N\left(\hat{\beta} = \left(X'X + n_\beta I_p\right)^{-1}\left(X'y + n_\beta \beta_0\right), \Sigma_\beta = \sigma^2\left(X'X + n_\beta I_p\right)^{-1}\right)$ and σ^2 follows an inverse gamma distribution such that $\sigma^2|\beta, X, y \sim IG(\alpha_*, \gamma_{**})$ and $\gamma_{**} = \left[(\beta - \hat{\beta})'\left(X'X + n_\beta I_p\right)(\beta - \hat{\beta}) + \phi\right]/2$. Since β and σ^2 are unobserved, we require a computational technique to estimate them from their posterior conditionals. In this case, we can use a Markov chain Monte Carlo (MCMC) technique, such as Gibbs sampling, or other computational techniques, such as the iterated conditional modes (ICM) algorithm to estimate β and σ^2.

3. Model 2: Real-Valued Parameters with an Unobserved Design Matrix

In this second model, we will describe the linear regression model with real-valued parameters and an unobserved design matrix. In this scenario, the only observed parameter is the dependent data y. The approach for estimating β, X, and σ^2 in this scenario requires a computational technique like Gibbs sampling.

3.1. Bayesian Approach to Linear Regression with Unobserved Design Matrix

Similar to the previous model, we first need to define the data likelihood, the prior distributions, and the posterior distribution before deriving the posterior marginals and conditionals. For the data likelihood, since the observed data y still follows a normal distribution, we can use Eq. 2.1 as the data likelihood for this model. For the prior distributions of β and σ^2, they will still follow normal and inverse gamma distributions as expressed in Eqs. 2.2 and 2.3, respectively. Since the design matrix X is also unobserved, it also requires a prior distribution to be placed on it. To keep consistency with conjugate priors, X will follow a normal distribution as shown in Eq. 3.1.

$$f(X|X_0, \sigma^2, n_x) \propto (\sigma^2)^{-\frac{np}{2}} \exp\left[-\frac{n_x}{2\sigma^2} tr\{(X - X_0)(X - X_0)'\}\right], \quad [3.1]$$

where tr is the trace of the $(X - X_0)(X - X_0)'$ matrix. Combining the data likelihood and prior distributions, the posterior distribution we obtain is

$$f(\beta, \sigma^2|X, y) \propto (\sigma^2)^{-\left(\frac{np+n+p+2\alpha}{2}+1\right)} \exp\left[-\frac{h_*}{2\sigma^2}\right], \qquad [3.2]$$

where $h_* = (y - X\beta)'(y - X\beta) + n_\beta(\beta - \beta_0)'(\beta - \beta_0) + n_x tr\{(X - X_0)(X - X_0)'\} + 2\gamma$.

3.2. Posterior Marginals and Conditionals

With the addition of X being an unobserved variable, we can no longer properly integrate out both X and σ^2 to estimate β. We can still integrate out σ^2 for both β and X, yielding

$$f(\beta|X,y) \propto \left[1 + \frac{1}{\nu_\beta}(\beta - \hat{\beta})' \left[\frac{\nu_\beta(X\prime X + n_\beta I_p)}{\phi}\right](\beta - \hat{\beta})\right]^{-\frac{\nu_\beta + p}{2}}, \qquad [3.3]$$

$$f(X|\beta,y) \propto \left[1 + \frac{1}{\nu_x}(X - \hat{X}) \left[\frac{\nu_x(\beta\beta\prime + n_x I_p)}{\theta}\right](X - \hat{X})'\right]^{-\frac{\nu_x + np}{2}}, \qquad [3.4]$$

where β follows a multivariate location-scale *t* distributions such that $\beta|X,y \sim t(\nu_\beta, \hat{\beta}, T)$, $E(\beta|X,y) = \hat{\beta} = (X'X + n_\beta I_p)^{-1}(X'y + n_\beta\beta_0)$, $Var(\beta|X,y) = \frac{\nu_\beta}{\nu_\beta - 2}T$, and X follows a matrix location-scale *t* distribution such that $X|\beta,y \sim Mt(\nu_x, \hat{X}, \Delta)$, $E(X|\beta,y) = \hat{X} = (y\beta' + n_x X_0)(\beta\beta' + n_x I_p)^{-1}$, $Var(X|\beta,y) = \frac{\nu_x}{\nu_x - 2}\Delta$. Since these posteriors marginals still have unobserved parameters, a computational technique is necessary for estimating β and X.

Like Model 1, for the posterior conditionals, instead of integrating out β, X, or σ^2, we can retain the unobserved parameters, mathematically determine the posterior conditionals, and utilize a computational technique to estimate β, X, and σ^2. These yield

$$f(\beta|X,\sigma^2,y) \propto |\Sigma_\beta|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\beta - \hat{\beta})'(\Sigma_\beta)^{-1}(\beta - \hat{\beta})\right], \qquad [3.5]$$

$$f(X|\beta,\sigma^2,y) \propto |\Sigma_X|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}tr\{(X - \hat{X})(\Sigma_X)^{-1}(X - \hat{X})'\}\right], \quad [3.6]$$

$$f(\sigma^2|\beta,X,y) \propto (\sigma^2)^{-(\alpha_* + 1)} \exp\left[-\frac{\gamma_{**}}{\sigma^2}\right], \qquad [3.7]$$

where β follows a multivariate normal distribution such that $\beta|X,\sigma^2,y \sim N\left(\hat{\beta} = (X'X + n_\beta I_p)^{-1}(X'y + n_\beta\beta_0), \Sigma_\beta = \sigma^2(X'X + n_\beta I_p)^{-1}\right)$, X follows a matrix normal distribution such that $X|\beta,\sigma^2,y \sim MN\left(\hat{X} = (y\beta' + n_x X_0)(\beta\beta' + n_x I_p)^{-1}, \Sigma_X = \sigma^2(\beta\beta' + n_x I_p)^{-1}\right)$, and σ^2 follows an inverse gamma distribution such that $\sigma^2|\beta,X,y \sim IG(\alpha_*, \gamma_{**})$, $\alpha_* = (np + n + p + 2\alpha)/2$, and $\gamma_{**} = [(y - X\beta)'(y - X\beta) + n_\beta(\beta - \beta_0)'(\beta - \beta_0) + n_x tr\{(X - X_0)(X - X_0)'\} + 2\gamma]/2$. Similar to the posterior marginals, the poster conditionals also require the use of a computational technique for estimating β, X, and σ^2. Since posterior marginals and conditionals for β are the location-scale *t* and normal distributions respectively, the expected values (mean) $\hat{\beta}$ and the mode are theoretically equivalent since the mean and the mode for the location-scale *t* and normal distributions are equal. This means that you can use the posterior marginals as a method of confirming the estimates from using the posterior conditionals (or vice versa). This can similarly be said for X.

4. Model 3: Complex-Valued Parameters with an Unobserved Design Matrix

In this third model, we will evaluate the linear regression model with complex-valued parameters and an unobserved design matrix. Similar to Model 2, the only observed parameter is the dependent data y and the approach for estimating β, X, and σ^2 also requires a computational method. For this model, we will introduce a real-valued isomorphic representation for the model.

4.1. Real-valued Isomorphic Representation

Transforming the complex-valued parameters into a real-valued representation allows for easier parameter estimation compared to the complex-valued counterpart. The complex-valued linear regression can be expressed as

$$y_c = X_c\beta_c + \varepsilon_c, \qquad [4.1]$$

with "*c*" subscripts representing that the parameter is complex-valued. For the real-valued isomorphic representation, each complex-valued parameter will be split into the real component (represented with an "*R*" subscript) and the imaginary component (represented with an "*I*" subscript). We can then express Eq 4.1 as

$$\begin{bmatrix} y_R \\ y_I \end{bmatrix} = \begin{bmatrix} x_R & -x_I \\ x_I & x_R \end{bmatrix} \begin{bmatrix} \beta_R \\ \beta_I \end{bmatrix} + \begin{bmatrix} \varepsilon_R \\ \varepsilon_I \end{bmatrix}, \; [4.2]$$

where $(\varepsilon_R, \varepsilon_I)' \sim N(0, \sigma^2 I_{2n})$. This linear model can be rewritten as Eq 1.2 where $y \in \mathbb{R}^{2n\times 1}$, $X \in \mathbb{R}^{2n\times 2p}$, $\beta \in \mathbb{R}^{2p\times 1}$, and $\varepsilon \in \mathbb{R}^{2n\times 1}$. We note that estimating $\hat{\beta}$ via Eq. 1.3 with an observed design matrix X_c yields similar results when the parameters are kept in complex-valued form compared with using its real-valued isomorphic representation, as exhibited in the Appendix.

4.2. Bayesian Approach using the Isomorphic Representation

For the real-valued isomorphic model, there are two representations for the design matrix X. The first one, expressed as X, is the skew-symmetric form expressed in Eq. 4.2. The second representation is expressed as $H = [x_R \quad x_I]$ yielding $H \in \mathbb{R}^{n\times 2p}$. This second representation is utilized to estimate the real and imaginary components of the design matrix only once since each component is expressed only one time in H, but twice in X. After estimating the real and imaginary components of the design matrix using the H representation, these components are then formulated into the X representation for estimating β and σ^2.

With this real-valued isomorphic representation, despite similarity to Model 2, we still need to define the data likelihood and prior distributions to obtain our posterior distribution. For the observed data, the real and imaginary components follow a normal distribution, yielding

$$f(y|X,\beta,\sigma^2) \propto (\sigma^2)^{-\frac{2n}{2}} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right]. \quad [4.3]$$

As for the prior distributions, β, X, and σ^2 will follow a multivariate normal distribution, matrix normal distribution, and an inverse gamma distribution, respectively, expressed as,

$$f(\beta|\beta_0,\sigma^2,n_\beta) \propto (\sigma^2)^{-\frac{2p}{2}} \exp\left[-\frac{n_\beta}{2\sigma^2}(\beta - \beta_0)'(\beta - \beta_0)\right], \quad [4.4]$$

$$f(H|H_0,\sigma^2,n_x) \propto (\sigma^2)^{-\frac{2np}{2}} \exp\left[-\frac{n_x}{2\sigma^2} tr\{(H - H_0)(H - H_0)'\}\right], \quad [4.5]$$

$$f(\sigma^2|\alpha,\gamma) \propto (\sigma^2)^{-(\alpha+1)} \exp\left[-\frac{\gamma}{\sigma^2}\right]. \quad [4.6]$$

One notable difference between this model and Model 2 with the data likelihood and the priors is the exponent of the σ^2 component of each equation (except Eq. 4.6) due to the dimensions of the parameters from the isomorphic representation. The resulting combination of the likelihood and the prior distributions is

$$f(\beta,\sigma^2|X,y) \propto (\sigma^2)^{-\left(\frac{2np+2n+2p+2\alpha}{2}+1\right)} \exp\left[-\frac{h_{**}}{2\sigma^2}\right], \quad [4.7]$$

where $h_{**} = (y - X\beta)'(y - X\beta) + n_\beta(\beta - \beta_0)'(\beta - \beta_0) + n_x tr\{(H - H_0)(H - H_0)'\} + 2\gamma$.

4.3. Posterior Marginals and Conditionals

Similar to Model 2, we cannot properly integrate out both X and σ^2 to estimate β, but can still integrate out σ^2 for both β and X to obtain the posterior marginals. These posterior marginals are

$$f(\beta|X,y) \propto \left[1 + \frac{1}{v_\beta}(\beta - \hat{\beta})'\left[\frac{v_\beta(X'X + n_\beta I_{2p})}{\phi}\right](\beta - \hat{\beta})\right]^{-\frac{v_\beta+2p}{2}}, \quad [4.8]$$

$$f(H|\beta,y) \propto \left|I_n + \frac{1}{v_x}(H - \hat{x})\left[\frac{v_x(CC' + n_x I_{2p})}{\theta}\right](H - \widehat{H})'\right|^{-\frac{v_x+2np}{2}}, \quad [4.9]$$

where β follows a multivariate location-scale *t* distribution such that $\beta|X,y \sim t(\nu_\beta, \hat{\beta}, T)$, $E(\beta|X,y) = \hat{\beta} = \left(X'X + n_\beta I_{2p}\right)^{-1}\left(X'y + n_\beta \beta_0\right)$, $Var(\beta|X,y) = \frac{\nu_\beta}{\nu_\beta - 2}T$, and X follows a matrix location-scale *t* distribution such that $H|\beta, y \sim Mt(\nu_x, \widehat{H}, \Delta)$, $E(H|\beta,y) = \widehat{H} = (YC' + n_x x_0)\left(CC' + n_x I_{2p}\right)^{-1}$, $Var(H|\beta,y) = \frac{\nu_x}{\nu_x - 2}\Delta$, $C = \begin{bmatrix} \beta_R & \beta_I \\ -\beta_I & \beta_R \end{bmatrix}$, and $Y = [y_R \quad y_I]$.

For the posterior conditionals for β, X, or σ^2, we have

$$f(\beta|X,\sigma^2,y) \propto \left|\Sigma_\beta\right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\beta - \hat{\beta})'\left(\Sigma_\beta\right)^{-1}\left(\beta - \hat{\beta}\right)\right], \qquad [4.10]$$

$$f(H|\beta,\sigma^2,y) \propto |\Sigma_X|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(H - \widehat{H}\right)(\Sigma_X)^{-1}\left(H - \widehat{H}\right)'\right], \qquad [4.11]$$

$$f(\sigma^2|\beta,X,y) \propto (\sigma^2)^{-(\alpha_*+1)} \exp\left[-\frac{\gamma_{**}}{\sigma^2}\right], \qquad [4.12]$$

where β follows a multivariate normal distribution such that $\beta|X,\sigma^2,y \sim N\left(\hat{\beta} = \left(X'X + n_\beta I_p\right)^{-1}\left(X'y + n_\beta \beta_0\right), \Sigma_\beta = \sigma^2\left(X'X + n_\beta I_p\right)^{-1}\right)$, H follows a matrix normal distribution such that $H|\beta,\sigma^2,y \sim MN\left(\widehat{H} = (YC' + n_x H_0)\left(CC' + n_x I_{2p}\right)^{-1}, \Sigma_X = \sigma^2\left(CC' + n_x I_{2p}\right)^{-1}\right)$, and σ^2 follows an inverse gamma distribution such that $\sigma^2|\beta,X,y \sim IG(\alpha_*, \gamma_{**})$, $\alpha_* = np + n + p + \alpha$, and $\gamma_{**} = \left[(y - X\beta)'(y - X\beta) + n_\beta(\beta - \beta_0)'(\beta - \beta_0) + n_x tr\{(H - H_0)(H - H_0)'\} + 2\gamma\right]/2$. Similar to Model 2, the posterior marginals and conditionals require a computational technique, like a Gibbs sampler or the ICM algorithm, for estimating β, X, and σ^2. The process for the Gibbs sampler and the ICM algorithm is outlined in the Appendix for Model 3. Also, like Model 2, we can use the posterior marginals of β and X to theoretically validate the results of the distributional mean or mode of their respective posterior conditionals.

4.4. Computational Techniques for Parameter Estimation

As discussed in the Posterior Marginals and Conditionals subsections, different computational strategies are either optional (Model 1) or required (Models 2 and 3) for posterior inference. In **Figure 1**, we illustrate the process of the Gibbs sampler and the ICM algorithm for Model 3, formulated using the real-valued isomorphic representation of the latent regression model. Because the computational procedures for Models 1 and 2 follow analogous steps, we focus exclusively on Model 3 as a representative case. The methods shown can also be implemented for the posterior marginals.

Before applying the computational methods, prior data is utilized to compute estimates for β_0, H_0, σ_0^2, n_β, n_x, α, and γ. For β_0, H_0, and σ_0^2, these are determined by estimating the regression coefficients, design matrix, and sample variance, respectively, from fully observed prior data, not the current observed data. The scalar hyperparameters n_β and n_x are set to be n_0, which is the number of samples utilized to estimate β_0 and H_0. We set the inverse-gamma hyperparameters to $\alpha = n_0 - 1$ and $\lambda = (n_0 - 1)\sigma_0^2$, which ensures that the prior mean of the variance equals σ_0^2 and allows n_0 to be interpreted as an effective prior sample size, a standard choice in conjugate Bayesian models for variance parameters[21].

Once the hyperpriors are computed and set initial values for $\beta_{(0)}$, $H_{(0)}$, and $\sigma_{(0)}^2$, we can start the computational process using either the Gibbs sampler or ICM algorithm as outlined in **Figure 1**. The exact distributions for the Gibbs sampler and modes for the ICM algorithm for each parameter in Model 3 are exhibited in **Table 1**. For the Gibbs sampler, there are a fixed set of burn-in iterations that are removed from the parameter estimation analysis since each sampling iteration is utilized for estimating mean, variance, confidence intervals, etc. of each parameter. For the ICM algorithm, there are no burn-in iterations since the algorithm only takes the last iteration as the mode estimate for each parameter. Both algorithms are completed once convergence is reached.

Both the Gibbs sampler and ICM algorithm provide notable advantages compared to the other. The Gibbs sampler provides a full sampling distribution for each parameter which allows for extensive statistical analysis. However, running a full Gibbs sampler can be computationally expensive, especially with a larger dataset and the number of iterations it may take to achieve convergence (can be up to hundreds of thousands of iterations). If we are only interested in the expected value (mean) of the sampling distribution for the regression coefficients β and design matrix X provided from the Gibbs sampler, we can simply use the ICM algorithm instead since the means and the modes of their posterior conditionals and marginals (normal and location-scale *t* distributions, respectively) are theoretically equivalent. This would save a significant amount of computational expense since the ICM algorithm can converge in as little as three iterations.

5. Discussion

In this review, we examined how unobserved design matrices and complex-valued parameters fundamentally alter the structure and estimation of linear regression models. By progressively moving from a classical real-valued regression framework with an observed design matrix

(Model 1) to increasingly realistic settings involving latent design matrices (Model 2) and complex-valued parameters (Model 3), we highlighted both the methodological challenges and the corresponding Bayesian solutions required for valid inference. In particular, we showed that while closed-form estimators are available in the simplest setting, Models 2 and 3 necessarily require computational techniques to jointly estimate the regression coefficients, latent design matrix, and residual variance.

A central theme of this work is the equivalence and interplay between posterior marginals and posterior conditionals in latent regression settings. For both the regression coefficients and the latent design matrix in Models 2 and 3, either formulation may be used for estimation, with posterior marginals providing theoretical validation for estimates obtained via posterior conditionals. This duality is especially valuable in practice, as it allows researchers to balance computational efficiency against inferential richness. When full uncertainty quantification is required, sampling-based methods such as the MCMC Gibbs sampler provide access to complete posterior distributions. When point estimation is sufficient, optimization-based alternatives such as the ICM algorithm offer substantial computational savings.

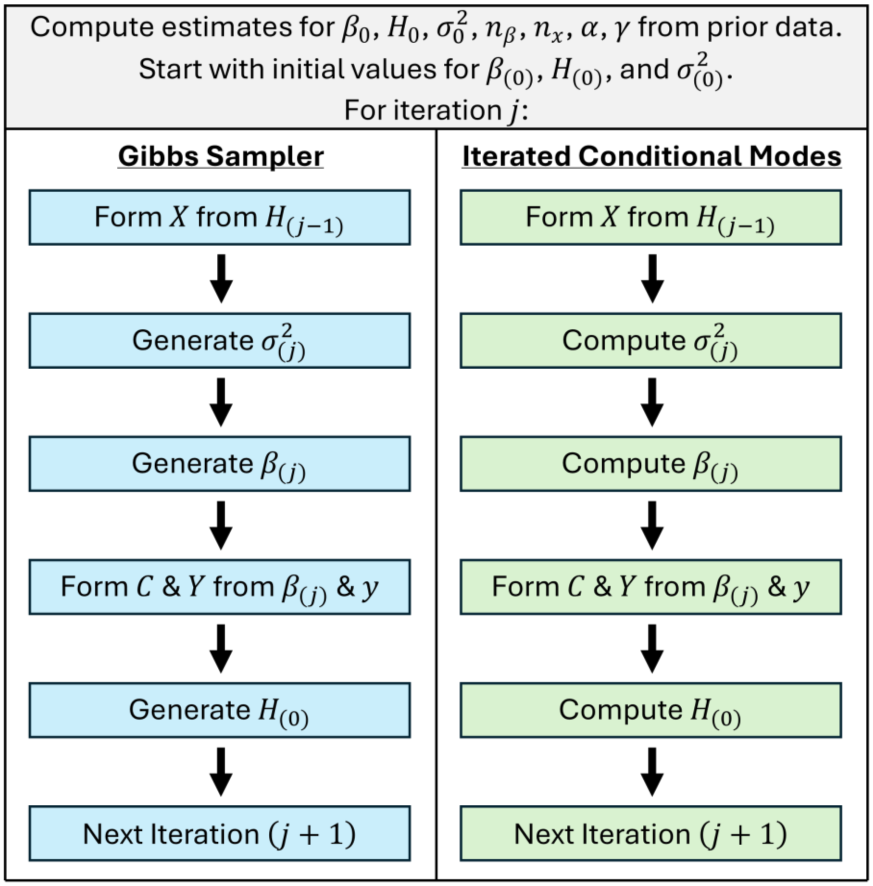
As emphasized in the Introduction, complex-valued latent regression models arise naturally across a wide range of scientific and engineering applications. Functional magnetic resonance imaging (fMRI) provides a particularly compelling example. The BSENSE[22] method operates in the image domain and employs a complex-valued latent regression formulation directly on the images, while BGRAPPA[23] applies the same model in the spatial-frequency domain with both methods improve reconstruction accuracy, enhance signal-to-noise ratio, and brain activity detection power. These applications demonstrate that explicitly modeling both latent structure and complex-valued data is not merely a theoretical extension, but a practical necessity for achieving state-of-the-art performance in modern imaging pipelines.

Several important directions for future work naturally follow from the framework developed here. First, all models considered in this review assume independence among observations and among regression coefficients, including independence between real and imaginary components. While this assumption simplifies analysis and computation, it may be restrictive in applications where temporal, spatial, or structural dependencies are present. Extending the latent regression framework to incorporate covariance structures among observations, regression coefficients, or both would allow the model to capture additional sources of dependence and potentially improve estimation accuracy and interpretability. More broadly, the

Bayesian formulation presented in this review provides a flexible foundation for extending latent regression models beyond the settings considered here.

Figures

Figure 1: Process for Model 3 parameter estimation via Gibbs sampling and ICM algorithm



Tables

Table 1: Model 3's parameter estimation from the posterior conditional distributions via Gibbs sampling and the ICM algorithm. The j subscript indicates the iteration number where $j = 1, \ldots, n$

Parameter	**Technique**	**Parameter Estimation from Posterior Conditional**
σ^2	Gibbs sampler	$\sigma^2_{(j)} \mid \beta_{(j-1)}, H_{(j-1)}, X, y \sim IG(\boldsymbol{\alpha}_*, \boldsymbol{\gamma}_{**})$ $\boldsymbol{\alpha}_* = np + n + p + \alpha$ $\boldsymbol{\gamma}_{**} = \big[\big(y - X\beta_{(j-1)}\big)'\big(y - X\beta_{(j-1)}\big) + n_\beta\big(\beta_{(j-1)} - \beta_0\big)'\big(\beta_{(j-1)} - \beta_0\big) + n_x tr\big\{\big(H_{(j-1)} - H_0\big)\big(H_{(j-1)} - H_0\big)'\big\} + 2\gamma\big]/2$
	ICM	$\sigma^2_{(j)} = \boldsymbol{\gamma}_{**}/(\boldsymbol{\alpha}_* + 1)$ $\boldsymbol{\alpha}_* = np + n + p + \alpha$ $\boldsymbol{\gamma}_{**} = \big[\big(y - X\beta_{(j-1)}\big)'\big(y - X\beta_{(j-1)}\big) + n_\beta\big(\beta_{(j-1)} - \beta_0\big)'\big(\beta_{(j-1)} - \beta_0\big) + n_x tr\big\{\big(H_{(j-1)} - H_0\big)\big(H_{(j-1)} - H_0\big)'\big\} + 2\gamma\big]/2$
β	Gibbs sampler	$\beta_{(j)} \mid X, \sigma^2_{(j)}, y \sim N\left(\left(X'X + n_\beta I_p\right)^{-1}\left(X'y + n_\beta \beta_0\right), \Sigma_\beta = \sigma^2_{(j)}\left(X'X + n_\beta I_p\right)^{-1}\right)$
	ICM	$\beta_{(j)} = \left(X'X + n_\beta I_p\right)^{-1}\left(X'y + n_\beta \beta_0\right)$
H	Gibbs sampler	$H_{(j)} \mid C, \sigma^2_{(j)}, Y \sim MN\left((YC' + n_x x_0)\left(CC' + n_x I_{2p}\right)^{-1}, \sigma^2_{(j)}\left(CC' + n_x I_{2p}\right)^{-1}\right)$
	ICM	$H_{(j)} = (YC' + n_x x_0)\left(CC' + n_x I_{2p}\right)^{-1}$

References

1. Petropulu, A. P., Zhang, R., & Lin, R. (2004). Blind OFDM Channel Estimation Through Simple Linear Precoding. IEEE Transactions on Wireless Communications, 3(2), 647–655.
2. Tugnait, J. K., & Li, T.-T. (2001). A Multistep Linear Prediction Approach to Blind Asynchronous CDMA Channel Estimation and Equalization. IEEE Journal on Selected Areas in Communications, 19(6), 1090–1102.
3. Schmidt, R. O. (1986). Multiple Emitter Location and Signal Parameter Estimation. IEEE Transactions on Antennas and Propagation, 34(3), 276–280.
4. Roy, R., & Kailath, T. (1989). ESPRIT—Estimation of Signal Parameters via Rotational Invariance Techniques. IEEE Transactions on Acoustics, Speech, and Signal Processing, 37(7), 984–995.
5. Comon, P. (1994). Independent Component Analysis, a New Concept? Signal Processing, 36(3), 287–314.
6. Yellin, D., & Weinstein, E. (1996). Multichannel Signal Separation: Methods and Analysis. IEEE Transactions on Signal Processing, 44(1), 106–118.
7. Cardoso, J. F., & Laheld, B. H. (1996). Equivariant Adaptive Source Separation. IEEE Transactions on Signal Processing, 44(12), 3017–3030.
8. Swami, A., Giannakis, G., & Shamsunder, S. (1994). Multichannel ARMA Processes. IEEE Transactions on Signal Processing, 42(4), 898–913.
9. Bell, A. J., & Sejnowski, T. J. (1995). An Information-Maximization Approach to Blind Separation and Blind Deconvolution. Neural Computation, 7(6), 1129–1159.
10. Cichocki, A., Unbehauen, R., & Rummert, E. (1994). Robust Learning Algorithm for Blind Separation of Signals. Electronics Letters, 30(17), 1386–1387.
11. Roth, Z., & Baram, Y. (1996). Multidimensional Density Shaping by Sigmoids. IEEE Transactions on Neural Networks, 7(5), 1291–1298.
12. Cao, X. R., & Liu, R. W. (1996). General Approach to Blind Source Separation. IEEE Transactions on Signal Processing, 44(3), 562–571.
13. Schreier, P. J., & Scharf, L. L. (2010). Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals. Cambridge University Press.
14. Lee, T. W., Bell, A. J., & Orglmeister, R. (1997). Blind Source Separation of Real World Signals. In Proceedings of the International Conference on Neural Networks, Vol. 4, 2129–2134.
15. Loizou, P. C. (2013). Speech Enhancement: Theory and Practice (2nd ed.). CRC Press.
16. Williamson, D. S., Wang, Y., & Wang, D. (2013). A Sparse Representation Approach for Perceptual Quality Improvement of Separated Speech. In Proc. IEEE ICASSP, 7015–7019.
17. Williamson, D. S., Wang, Y., & Wang, D. (2014). A Two-Stage Approach for Improving the Perceptual Quality of Separated Speech. In Proc. IEEE ICASSP, 7034–7038.
18. Gerchberg, R. W., & Saxton, W. O. (1972). A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures. Optik, 35, 237–246.
19. Fienup, J. R. (1982). Phase Retrieval Algorithms: A Comparison. Applied Optics, 21(15), 2758–2769.
20. Mansour, H., Liu, D., Kamilov, U., Boufounos, P.T., Sparse Blind Deconvolution for Distributed Radar Autofocus Imaging, IEEE Transactions on Computational Imaging, DOI: 10.1109/TCI.2018.2875375, Vol. 4, No. 4, pp. 537-551, December 2018.
21. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian Data Analysis (3rd ed.). Chapman & Hall/CRC.

22. Chase J Sakitis, D Andrew Brown, Daniel B Rowe, A Bayesian complex-valued latent variable model applied to functional magnetic resonance imaging, *Journal of the Royal Statistical Society Series C: Applied Statistics*, Volume 74, Issue 1, January 2025, Pages 100–125.
23. Chase J. Sakitis. Daniel B. Rowe. "A Bayesian approach to GRAPPA parallel fMRI image reconstruction increases SNR and power of task detection." Ann. Appl. Stat. 19 (2) 1473 - 1493, June 2025. https://doi.org/10.1214/24-AOAS1962

Appendix

Below exhibits the math showing that the complex, ordinary least squares estimate for β is equal if left in complex-valued form compared to the real-valued isomorphic representation when the design matrix is known.

$$\beta_c = \beta_R + i\beta_I \qquad X_c = X_R + iX_I \qquad y_c = y_R + iy_I$$

$$\widehat{\beta_c} = (X_c^H X_c)^{-1}(X_c^H y_c) \qquad \text{(complex-valued OLS)}$$

$$\widehat{\beta_c} = [(X_R + iX_I)^H (X_R + iX_I)]^{-1}[(X_R + iX_I)^H (y_R + iy_I)] \qquad \text{(R/I components)}$$

$$\widehat{\beta_c} = [(X_R^T - iX_I^T)(X_R + iX_I)]^{-1}[(X_R^T - iX_I^T)(y_R + iy_I)] \qquad \text{(distribute Hermitian)}$$

$$\widehat{\beta_c} = [X_R^T X_R + iX_R^T X_I - iX_I^T X_R - i^2 X_I^T X_I]^{-1}[X_R^T y_R + iX_R^T y_I - iX_I^T y_R - i^2 X_I^T y_I] \qquad \text{(multiply parentheses)}$$

$$\widehat{\beta_c} = [X_R^T X_R + iX_R^T X_I - iX_I^T X_R + X_I^T X_I]^{-1}[X_R^T y_R + iX_R^T y_I - iX_I^T y_R + X_I^T y_I] \qquad \text{(simplified } i^2\text{)}$$

$$\widehat{\beta_c} = [(X_R^T X_R + X_I^T X_I) + i(X_R^T X_I - X_I^T X_R)]^{-1}[(X_R^T y_I + X_R^T y_I) + i(X_R^T y_I - X_I^T y_R)] \qquad \text{(rearrange)}$$

$$A = (X_R^T X_R + X_I^T X_I) \qquad B = (X_R^T X_I - X_I^T X_R) \qquad C = (X_R^T y_I + X_R^T y_I) \qquad D = (X_R^T y_I - X_I^T y_R)$$

$$\widehat{\beta_c} = (A + iB)^{-1}(C + iD) \qquad \text{(replace equations with letters)}$$

$$(A + iB)\widehat{\beta_c} = C + iD \qquad \text{(multiply both sides by } (A + iB)\text{)}$$

$$(A + iB)(\beta_R + i\beta_I) = C + iD \qquad \text{(express } \beta \text{ in complex form)}$$

$$A\beta_R + iA\beta_I + iB\beta_R + i^2 B\beta_I = C + iD \qquad \text{(multiply parentheses)}$$

$$(A\beta_R - B\beta_I) + i(A\beta_I + B\beta_R) = C + iD \qquad \text{(rearrange)}$$

$$\begin{bmatrix} A\beta_R - B\beta_I = C \\ A\beta_I + B\beta_R = D \end{bmatrix} \qquad \text{(Write as system of equations)}$$

$$\begin{bmatrix} A & -B \\ B & A \end{bmatrix}\begin{bmatrix} \beta_R \\ \beta_I \end{bmatrix} = \begin{bmatrix} C \\ D \end{bmatrix} \qquad \text{(breakdown of the system into matrix/vectors)}$$

$$\begin{bmatrix} \beta_R \\ \beta_I \end{bmatrix} = \begin{bmatrix} A & -B \\ B & A \end{bmatrix}^{-1}\begin{bmatrix} C \\ D \end{bmatrix} \qquad \text{(divide matrix on both sides)}$$

$$\begin{bmatrix} \beta_R \\ \beta_I \end{bmatrix} = \begin{bmatrix} (X_R^T X_R + X_I^T X_I) & -(X_R^T X_I - X_I^T X_R) \\ (X_R^T X_I - X_I^T X_R) & (X_R^T X_R + X_I^T X_I) \end{bmatrix}^{-1}\begin{bmatrix} (X_R^T y_I + X_R^T y_I) \\ (X_R^T y_I - X_I^T y_R) \end{bmatrix} \qquad \text{(write out equations for each letter)}$$

$$\begin{bmatrix} \beta_R \\ \beta_I \end{bmatrix} = \begin{bmatrix} (X_R^T X_R + X_I^T X_I) & (X_I^T X_R - X_R^T X_I) \\ (X_R^T X_I - X_I^T X_R) & (X_R^T X_R + X_I^T X_I) \end{bmatrix}^{-1}\begin{bmatrix} (X_R^T y_I + X_R^T y_I) \\ (X_R^T y_I - X_I^T y_R) \end{bmatrix} \qquad \text{(distribute minus sign)}$$

$$\begin{bmatrix} \beta_R \\ \beta_I \end{bmatrix} = \left[\begin{bmatrix} X_R^T & X_I^T \\ -X_I^T & X_R^T \end{bmatrix}\begin{bmatrix} X_R & -X_I \\ X_I & X_R \end{bmatrix}\right]^{-1}\left[\begin{bmatrix} X_R^T & X_I^T \\ -X_I^T & X_R^T \end{bmatrix}\begin{bmatrix} y_R \\ y_I \end{bmatrix}\right] \qquad \text{(break up each component)}$$

$$\begin{bmatrix} \beta_R \\ \beta_I \end{bmatrix} = \left[\begin{bmatrix} X_R & -X_I \\ X_I & X_R \end{bmatrix}^T\begin{bmatrix} X_R & -X_I \\ X_I & X_R \end{bmatrix}\right]^{-1}\left[\begin{bmatrix} X_R & -X_I \\ X_I & X_R \end{bmatrix}^T\begin{bmatrix} y_R \\ y_I \end{bmatrix}\right] \qquad \text{(move transpose)}$$

$\beta = (X^T X)^{-1}(X^T y)$ (simplify)

$$\beta = \begin{bmatrix} \beta_R \\ \beta_I \end{bmatrix} \qquad X = \begin{bmatrix} X_R & -X_I \\ X_I & X_R \end{bmatrix} \qquad y = \begin{bmatrix} y_R \\ y_I \end{bmatrix}$$