

Understanding Coagulopathy using Multi-view Data in the Presence of Sub-Cohorts: A Hierarchical Subspace Approach

Arya A. Pourzanjani

*Department of Computer Science
University of California, Santa Barbara
Santa Barbara, CA, U.S.A.*

ARYA@ENGINEERING.UCSB.EDU

Tie Bo Wu

*Department of Mechanical Engineering
University of California, Santa Barbara
Santa Barbara, CA, U.S.A.*

TIEBO@UCSB.EDU

Richard M. Jiang

*Department of Computer Science
University of California, Santa Barbara
Santa Barbara, CA, U.S.A.*

RMJIANG@CS.UCSB.EDU

Mitchell J. Cohen

*Denver Health Medical Center
Denver, CO, U.S.A.*

Linda R. Petzold

*Department of Computer Science
University of California, Santa Barbara
Santa Barbara, CA, U.S.A.*

PETZOLD@ENGINEERING.UCSB.EDU

Abstract

Death from trauma is most often the result of uncontrollable bleeding as a result of Acute Traumatic Coagulopathy (ATC), a disease that manifests itself differently in different sub-cohorts of trauma patients. Understanding the mechanisms of ATC and how existing patient tests can inform us about these mechanisms is key to treating the disease. We introduce a hierarchical Canonical Correlation Analysis (CCA) model that captures a lower dimensional representation of the coagulation system based on blood protein and other tests. The hierarchical nature of the model is ideal in the setting where multiple sub-cohorts are present, but statistical strength can reasonably be borrowed from similar groups. We illustrate how the model may be useful in understanding and treating ATC.

1. Introduction

Trauma is the leading cause of death between the ages of 1 and 44 (Hoyert and Xu, 2012). Major trauma often induces a coagulopathic state known as Acute Traumatic Coagulopathy (ATC) that manifests in increased bleeding and resultant mortality (Brohi et al., 2003). Despite the coagulation cascade being a well-studied field in biochemistry, the causes and mechanisms behind ATC are still uncertain (Hess et al., 2008).

One of the primary difficulties in treating ATC is information scarcity. During trauma care, the state of the coagulation system is often assessed by thromboelastography (TEG). This assay allows for quick assessment of the patient’s clotting potential using the viscoelastic properties of the patient’s blood such as time to clot formation, clotting rate, clot strength, and rate of clot lysis. Unfortunately, TEG does not provide much insight on specific clotting factor levels which can highlight the abnormalities causing ATC. Assays that test for plasma protein levels can take much longer than TEG or may not be available. Thus, treatment decisions must be made with limited information. If we can use historical data and other patient data to ascertain the relationship between plasma protein levels and TEG, decision making can be improved at the point of care.

While some protein data has been collected and analyzed (Kutcher et al., 2013), a consensus mechanism for ATC remains to be found. One possible reason for this is that there may not be a single mechanism behind ATC. Currently, ATC is characterized primarily by clot times, but assuming a single mechanism can obfuscate the more likely heterogeneous nature of the disease that has been alluded to previously in the literature (Kutcher et al., 2013). The coagulation system is a complex network of reactions with many positive and negative feedback mechanisms (Gonzalez et al., 2014). Because of this, there are many possible mechanisms that can cause the system to fail. Distilling these mechanisms into a small collection of easily-interpretable variables would go a long way towards better understanding of ATC. This could potentially enable recognition of phenotypes using limited information such as TEG results, in order to inform decision making in practice.

Further complicating the issue is the inherent heterogeneity in patients and injury types. Coagulopathy is not relegated to any particular form of trauma or injury mechanism, e.g. car accidents or gunshot wounds. A-priori, we would expect that patients with different types of injury might have different coagulopathic phenotypes and cannot be grouped together in an analysis.

We propose a probabilistic graphical model that extends canonical correlation analysis (CCA) to distill the complex network of interactions in trauma-induced coagulopathy into a small set of easy-to-understand variables. The model extends CCA in such a way as to allow different sub-cohorts (based on injury mechanism) to be represented by different latent variables, i.e. subspaces. At the same time, the model is able to “borrow” information between similar sub-cohorts when data in a particular sub-cohort is sparse. This latter feature is accomplished by placing a hierarchical prior over the orthornormal matrices of weights used in CCA, via the recently introduced Givens transform (Pourzanjani et al., in preparation). To our knowledge this is the first application of hierarchical modeling (in the Bayesian sense) to CCA, PCA, or any kind of model with orthornormal matrix parameters, as this task was previously intractable for even small problems without a change of representation that the Givens transform provides.

We show how, when applied to blood protein assays and TEG measurements, our model can find different low dimensional descriptions of the coagulation system for different sub-cohorts of injured patients, and how the model can share information across sub-cohorts using hierarchical modeling. We then illustrate how the shared subspace properties of CCA may be used in a clinical setting to phenotype patients and guide treatment using TEG measurements (i.e. without expensive and time-consuming protein assays).

2. Cohort

Our dataset consists of whole blood collected from 174 patients upon arrival to the emergency department of a Level I trauma center between 2005 to 2015 as part of an ongoing study.

2.1 Cohort and Sub-Cohort Selection

To remove any bias arising from age and differing injury severity we selected patients below the age of 45 and with an injury severity score greater than or equal to 25, a common threshold used to mark severe injury in past studies. We examined patients suffering from gun shot wounds (GSW), motor vehicle collisions (MVC), stab wounds (SW) and assault, representing a diverse set of possible injury mechanisms. These divisions are also representative of groups with larger sample sizes (GSW) and smaller sample sizes (assault) to illustrate the inferential power of our hierarchical model. A summary of our sub-cohorts is summarized in table 2.

2.2 Feature Choices

For each patient, we examined the TEG and blood protein assays. These consist of values of R, K, MA, and Ly30 for TEG and FactorII, FactorX, Protein C, D-Dimer, Fibrinogen, and Platelets for blood protein. The values are collectively referred to as x_{teg} and x_{prot} for the remainder of this work. Table 3 in the appendix summarizes the number of missing entries for these values amongst the 174 total patients. We describe in the methods section how we account for these missing values.

3. Methods

We start with a description of how we handle missing values in our analysis and give a short introduction into Canonical Correlation Analysis. Next, we provide a description of how we conduct full Bayesian posterior inference over the orthornormal matrix parameters present in CCA using the Givens transform and Stan (Carpenter et al., 2016). Finally we describe our novel extension to CCA that takes into account the heterogeneity between sub-cohorts present in our dataset via Bayesian hierarchical modeling.

3.1 Missing Values

We use multiple imputation as described in (Van Buuren, 2012) to conduct inference while properly accounting for the extra uncertainty that arises from missing data. Specifically, we draw five imputed datasets, conduct Hamiltonian Monte Carlo (HMC) inference in Stan separately on these imputed datasets, and then combine all posterior samples (after sampling is complete) in to a single pool of samples. In this way our posterior samples represent a mixture of posteriors under various imputation possibilities. For actual imputation we use a nonparametric random imputation method based on CART described in (Van Buuren, 2012).

3.2 CCA and Probabilistic CCA (PCCA)

Canonical correlation analysis (CCA) is an extension of PCA used to extract cross-covariance information about a pair of related datasets, or views, that are tied to a common set of samples (Murphy, 2012). The classical formulation of CCA finds a pair of linear projections that maximize the shared variance between the two views. Inspecting the obtained projections allows one to observe the level of commonality across views.

The probabilistic formulation of CCA (PCCA), (Bach and Jordan, 2005), posits a generative process for the same task. Formally, if we have two sets of different data types, or views, $x_{\text{prot}} \in \mathbb{R}^{D_{\text{prot}}}$ and $x_{\text{teg}} \in \mathbb{R}^{D_{\text{teg}}}$, we can define for each view a set of orthonormal loading matrices W and B and corresponding diagonal matrices, Λ and Γ , that describe the variance explained by each latent dimension. Then for each sample $i \in (1, \dots, N)$, we can define three lower-dimensional latent variables; one for each view, $z_{\text{prot}}^{(i)} \in \mathbb{R}^{L_{\text{prot}}}$ and $z_{\text{teg}}^{(i)} \in \mathbb{R}^{L_{\text{teg}}}$, and a shared latent variable that connects both views, $z_s^{(i)} \in \mathbb{R}^{L_s}$. We define $z^{(i)} := (z_{\text{prot}}^{(i)}, z_{\text{teg}}^{(i)}, z_s^{(i)})$. A single data point for sample i is generated according to the following specification:

$$z^{(i)} \sim \mathcal{N}(z_{\text{prot}}^{(i)} | 0, I_{L_{\text{prot}}}) \mathcal{N}(z_s^{(i)} | 0, I_{L_s}) \mathcal{N}(z_{\text{teg}}^{(i)} | 0, I_{L_{\text{teg}}}) \quad (1)$$

$$x_{\text{prot}}^{(i)} \sim \mathcal{N}(x_{\text{prot}}^{(i)} | B_x \Gamma_x z_{\text{prot}}^{(i)} + W_{\text{prot}} \Lambda_{\text{prot}} z_s^{(i)} + \mu_{\text{prot}}, \sigma_{\text{prot}}^2 I_{D_{\text{prot}}}) \quad (2)$$

$$x_{\text{teg}}^{(i)} \sim \mathcal{N}(x_{\text{teg}}^{(i)} | B_y \Gamma_y z_{\text{teg}}^{(i)} + W_{\text{teg}} \Lambda_{\text{teg}} z_s^{(i)} + \mu_{\text{teg}}, \sigma_{\text{teg}}^2 I_{D_{\text{teg}}}) \quad (3)$$

The maximum likelihood estimate (MLE) for this model converges to the solution of classical CCA up to an invariant rotation of the axis as shown in (Bach and Jordan, 2005). The intuition under this generative process is that by introducing a shared latent variable z_s , the learned W matrices capture information that is shared between the two views while the B matrices capture information that is not contained in the other view. A graphical model of this is shown in Figure 1.

3.2.1 AUTOMATIC DIMENSIONALITY SELECTION

We note that the diagonal matrices Λ and Γ serve as importance weights, in the sense that they describe the weight given to each latent dimension in predicting the higher dimensional data. When these weights are close to zero, they indicate that a latent dimension has no predictive relationship to the data. Thus, examining these weights can yield insight into the inherent dimensionality of the data. Specifically, using full posterior draws of Λ and Γ allows us to make probabilistic statements about the inherent dimensionality of our data, e.g. if posterior draws show that a weight is below some value with high probability we can conclude that the latent dimension is unnecessary. We use this type of analysis for selection of the latent dimension in our results. We place Cauchy priors over importance weight as well as matrix coefficients to induce sparsity as described in the appendix.

3.3 Hierarchical Modeling and Hierarchical CCA

Treating dissimilar patients, such as patients with different types of injuries, as a single homogeneous group will lead to statistical bias in any inference. On the other hand, a

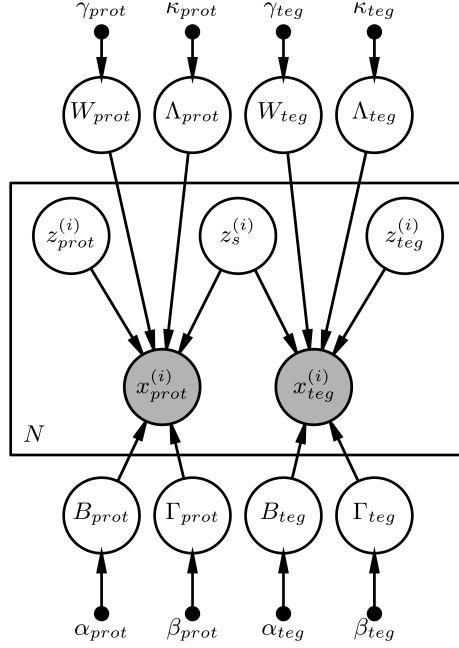


Figure 1: Graphical Model of a Non-Hierarchical CCA for our dataset. We attach various zero-centered Cauchy priors on the angles of the orthonormal matrices under the Givens parameterization to induce sparsity in our solutions.

full model where separate parameters are allocated and estimated for each of a multitude of sub-cohorts can lead to large models with many parameters that usually result in high uncertainty estimates in the fully Bayesian case, and overfit estimates in the maximum likelihood case (lest we provide an inordinate amount of data). Even in the age of big data, this can be an issue, as once we have appropriately sub-cohorted by all relevant features, e.g. injury type, sex, race, age, weight, etc. we are left with small sample sizes within each sub-cohort. Furthermore, certain sub-cohorts are prone to having less data, e.g. severely injured patients usually have fewer test values available because there was no time to collect data for those patients!

Hierarchical models are well known in Bayesian data analysis as an elegant solution to this conundrum. It proceeds by treating sub-cohorts as separate entities with their own respective parameters, but additionally modeling the dependence between the groups with a common prior distribution over the parameters (Gelman et al., 2014). Hierarchical models are very flexible because intuitively they can “adapt” to data by “shrinking” the value of similar groups to a common value when data in a group is sparse, but they allow parameter estimates to approach the value given by the data as more data is made available. The hierarchical model includes as a special case the “full” model where each group gets their own parameter that is estimated separately and a “null” model where each group is treated as one big homogeneous group. Hierarchical models allow us to have a model somewhere between these two extremes.

3.3.1 HIERARCHICAL CCA

In our probabilistic graphical model, we desire separate CCA parameters for each sub-cohort of injured patients. This calls for separate orthonormal matrices for each sub-cohort, and thus for hierarchical modeling, a prior distribution over orthonormal matrices. At least two distributions over orthonormal matrices exist in the statistics literature (Muirhead, 2009). One such distribution is the Matrix Langevin distribution. Incorporating these distributions into a probabilistic graphical model and conducting inference is difficult however, because evaluating their density functions involves computing the hypergeometric function of a matrix argument, a problem shown to be difficult even for small matrices (Koev and Edelman, 2006). We detail in the following section how we use the recently introduced Givens transform to build a hierarchical CCA model and perform joint analysis of the sub-cohorts.

3.4 Givens Transform for Graphical Models

Training of the model was done in Stan using Hamiltonian Monte Carlo (HMC) sampling of posteriors for all latent variables and unknown parameters. We provide a link to our Stan model file on Github.

In order to conduct full Bayesian inference on weight matrices and set hierarchical priors on such matrices, we require a way to sample on the space of orthonormal matrices, preferably using a robust sampling method like HMC, a difficult problem in general due to the constraint the samples of matrices must satisfy (Byrne and Girolami, 2013). While methods exist for HMC sampling of posteriors of general constrained parameters (Brubaker et al., 2012; Byrne and Girolami, 2013), these methods treat constrained and unconstrained parameters separately and require separate numerical integrators for each type of parameter, making them difficult to implement in larger probabilistic graphical models, with complicated priors, such as the model in described in the preceeding section.

In order to sample posteriors of orthonormal matrices, we instead appeal to the recently introduced Givens transform (Pourzanjani et al., "In preparation"). Rather than modifying the core HMC algorithm, the Givens transform represents an n, p orthonormal matrix by a product of $np - p(p + 1)/2$ rotation matrices, allowing us to use existing HMC implementations by sampling on the unconstrained space of angles. Making use of the Givens transform provides us with two distinct benefits over existing sampling methods on orthonormal matrices:

- Straightforward implementation in a probabilistic programming language such as Stan (Carpenter et al., 2016), allowing for simple expansion to large models.
- Placing complicated priors (e.g. hierarchical priors) over orthonormal matrices.

Because the Givens transform represents orthonormal matrices as angles, we can place hierarchical priors over orthonormal matrices in a straight-forward manner by simply placing priors over the angles. In our model we do just that, placing truncated normal priors over the respective angles of each group of injured patients. Our full probabilistic graphical model is the model described in Figure 1, but with the number of parameters multiplied by four, for the four different sub-cohorts described earlier, and the truncated normal prior that acts as a hierarchical prior over the orthonormal matrices.

4. Results

In the first sub-section we apply our model to plasma protein assays and TEG measurements to find low dimensional descriptions of the coagulation system for each respective sub-cohort of injured patients. We illustrate how in a probabilistic framework, we can assess with confidence the differences in sub-cohorts, as well as assess how much information our model captures, using posterior information. We then briefly describe the affect of our hierarchical inference.

In the second subsection we explain how the shared subspace properties of CCA may be used in a clinical setting to phenotype patients and guide treatment using TEG measurements, without protein assays.

4.1 Finding Distinct Phenotypes of Coagulopathy

Table 1 shows point estimates of the orthornormal matrices of weights, W_{prot} and W_{teg} , that relate the common latent variable, z_s , to measured data, x_{prot} and x_{teg} , for both the gun shot wound (GSW) and motor vehicle collision (MVC) sub-cohorts. Both groups of patients share commonalities in their latent variables. For example, their first latent variables correspond to high FactorII, FactorX, and platelets, but also to low TEG K (speed of clot formation). On the other hand, the model captures differences between the different groups, e.g. FactorX is a smaller component of the first latent variable in the GSW sub-cohort, than in the MVC sub-cohort. Full posterior inference provided by Stan and the Givens transform allows us to compare these relationships probabilistically using full posteriors rather than relying on point estimates (see Figure 2 for an example). From full posterior draws we can simply count the number of posterior draws where the FactorX weight parameter is higher in the GSW sub-cohort than the MVC sub-cohort, giving us a posterior probability of 0.6, in this case of the statement being true, given the data we observed.

Other differences between the two sub-cohorts are apparent in the point estimate of the weight matrix. The model reveals that the second latent variable in the GSW sub-cohort is tied strongly to low D-Dimer and low Ly30, while in the MVC sub-cohort the second latent variable is actually tied to high D-Dimer and low Ly30. Figure 3 visually illustrates this sub-cohort specific relationship our model was able to find.

In Figure 4 we compare posterior distributions of the “importance” parameters Λ and Γ of each of the latent dimensions. The plot can be suggestive of different inherent dimensionalities across the different subcohorts. For example the assault group has a posterior for the weight of the second shared dimension that is closer to zero while the GSW group’s weight for that same weight is concentrated more around a smaller positive value, possibly indicating that of the GSW victims there is an extra dimension to be kept track of in their shared latent space.

Lastly, we make note of the shrinkage properties of the hierarchical prior. We conducted separate, non-hierarchical inferences (not shown here) for the GSW group and MVC groups and found that the posterior distributions of their parameters, in particular the W_{prot} and W_{teg} matrices, were left relatively unchanged from the overall hierarchical estimates shown here. These groups had 85 and 52 patients in them respectively. On the other hand, we found much wider posteriors on parameters when estimating the assault group, which

Table 1: Point estimates for the orthonormal matrix of weights relating the common (between the protein and TEG data) latent variable to data for each respective sub-cohort. Point estimates were obtained by taking the sample with the largest log probability amongst all 20,000 posterior samples. Cells are colored by intensity of estimated values. Red indicates a strong positive correlation, light red a weak positive correlation, blue a strong negative correlation, and light blue a weak negative correlation.

	GSW		MVC	
Proteins	Latent 1	Latent 2	Latent 1	Latent 2
FactorII	0.47	-0.22	0.46	-0.38
FactorX	0.56	-0.09	0.82	0.05
ATIII	0.13	0.06	-0.02	0.57
D-Dimer	-0.15	-0.95	-0.05	0.51
Fibrinogen	0.10	0.17	0.22	0.30
Platelets	0.64	-0.02	0.26	0.41
TEG				
R	-0.74	0.06	0.00	-0.03
K	-0.42	-0.07	-0.84	0.14
MA	0.51	0.14	0.53	0.21
Ly30	-0.06	0.98	0.01	0.97

consisted of only 16 patients, where as the hierarchical estimate yields narrower posteriors that are closer to the estimates of the other sub-cohorts.

4.2 Using TEG and Injury Data to Predict Phenotypes

Intuitively, the hierarchical CCA model distills the information available in protein and TEG data into shared latent components and ties together the relationship between proteins and TEG by specifying how much of the information carried in these measurements can be ascertained simply by knowing the values of the latent variables. For example, in Table 1 we see that latent variable 1 of the GSW sub-cohort is tied to high FactorII and FactorX levels in the protein category, and low R, low K, and high MA in the TEG category. Thus intuitively, given values from a TEG test of a patient and their injury type, we should be able to back out a posterior distribution of the underlying protein values of that patient. In the case of a GSW victim with low R, low K, and high MA, the CCA model suggests that there is a good chance that this patient could have high FactorII and FactorX levels. This information on possible values of the underlying protein levels could serve as valuable information in a real life trauma setting where interventions are guided based on the state of the coagulation system. For example, plasma transfusions are administered when it is thought that a patient’s Factor levels are low. Figure 5 shows posterior draws of what a GSW patient’s FactorII will look like given a TEG reading with low R, low K, and high MA.

The diagonal weight matrix Λ can also roughly tell us how much coagulation protein information we can extract using TEG readings. High posterior values for the weights

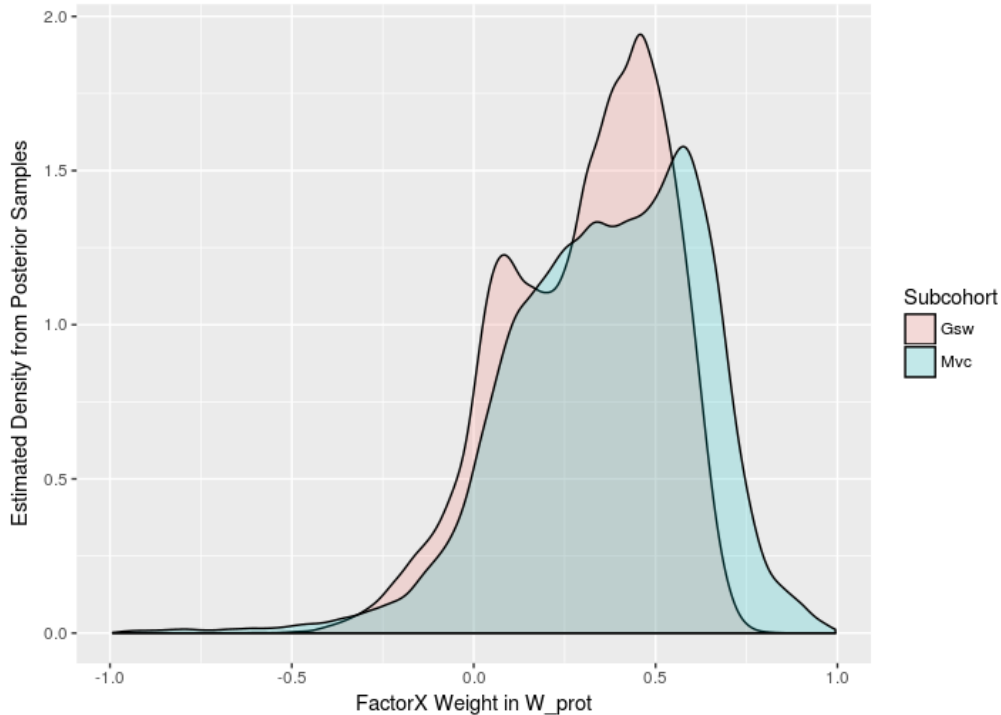


Figure 2: Densities estimated from posterior samples of the FactorX weight in the matrix W_{prot} , in the GSW and MVC sub-cohorts respectively. These posterior results indicate the MVC sub-cohort are represented by a latent dimension that places more weigh on FactorX.

connecting the shared latent variable to the protein data signify that variance in the protein data can be readily explained by information that is shared between both the protein and TEG views. Low values of those weights or high values of the Γ weights, connecting the latent protein variables to the protein data, would signify that there is information in the protein data that cannot be captured by the shared latent variable and thus can not be captured by TEG.

5. Discussion and Related Work

Our results show how trauma data can be distilled into low dimensional latent components in such a way as to respect the inherent difference in sub-cohorts of trauma patients. We show how differences in different groups can manifest themselves in different models, and how Bayesian posterior analysis of latent variable “weights” can be used to assess the inherent dimensionality of data in different groups. We then showed the benefits of hierarchical modeling as it pertained to our analysis, and why hierarchical modeling is an essential tool when sub-cohorting. We ended with a possible use case of how understanding the underlying relationship of coagulation proteins and TEG measurements via latent variables can help our understanding and the treatment of trauma. While we acknowledge that sub-cohorting by injury type may not be the optimal sub-cohort strategy, our method can be

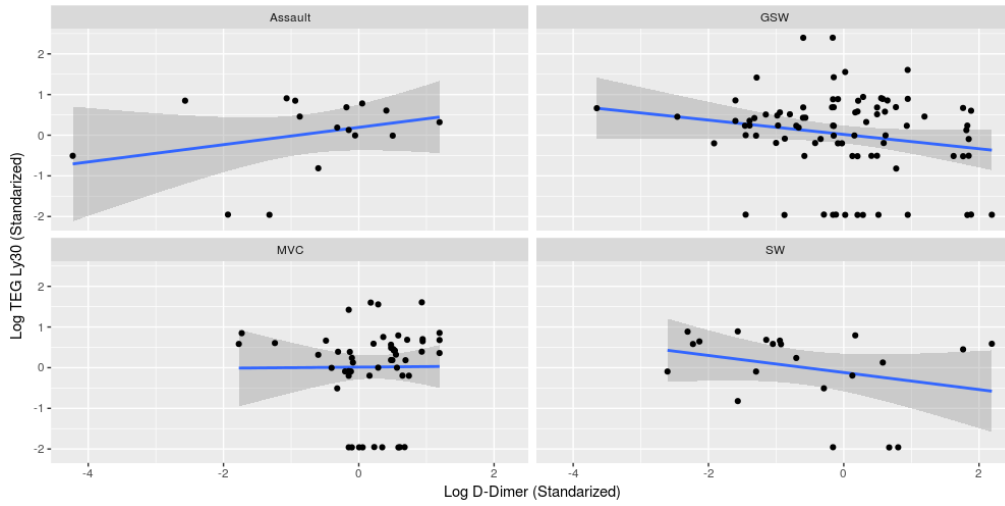


Figure 3: Relationship between D-Dimer and TEG Ly30 for different sub-cohorts of patients. For gun shot wounds victims and stab wound victims (both penetrating injuries) the two variables have a negative correlation where as for assault and motor vehicle collisions (both blunt injuries) the relationship is slightly positive.

used in more general settings and especially in medicine where sub-cohorting is important to analysis.

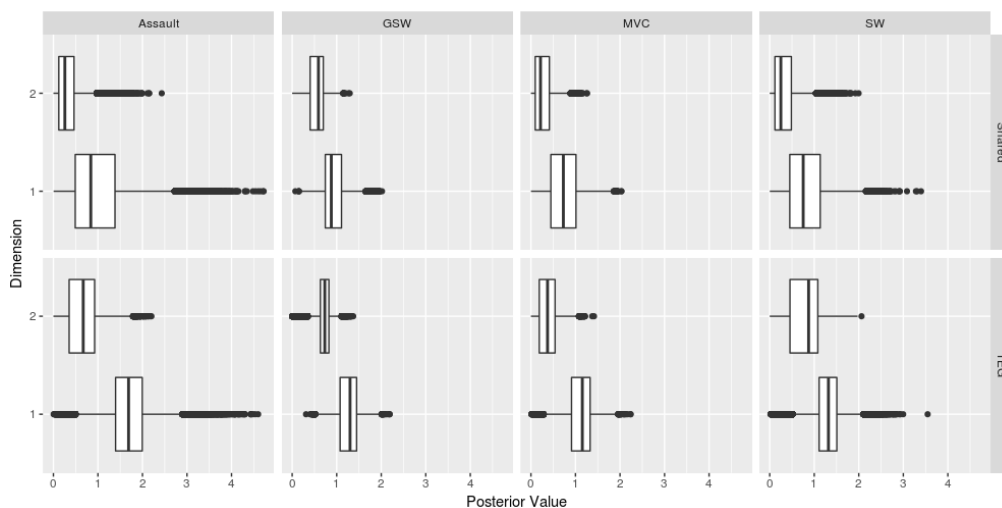


Figure 4: Posterior distributions of the “importance” parameters Λ and Γ of each of the latent dimensions. In each facet we show a boxplot for the importance parameters of latent dimension 1 and latent dimension 2, using samples from their posterior distributions. These weights describe the exploratory power the latent variables have in describing our data (see methods section). The row of shared facets represent importance weights for the latent variable that is shared between the protein and TEG views while the row of TEG facets shows posteriors of importance weights for the TEG-only latent variable.

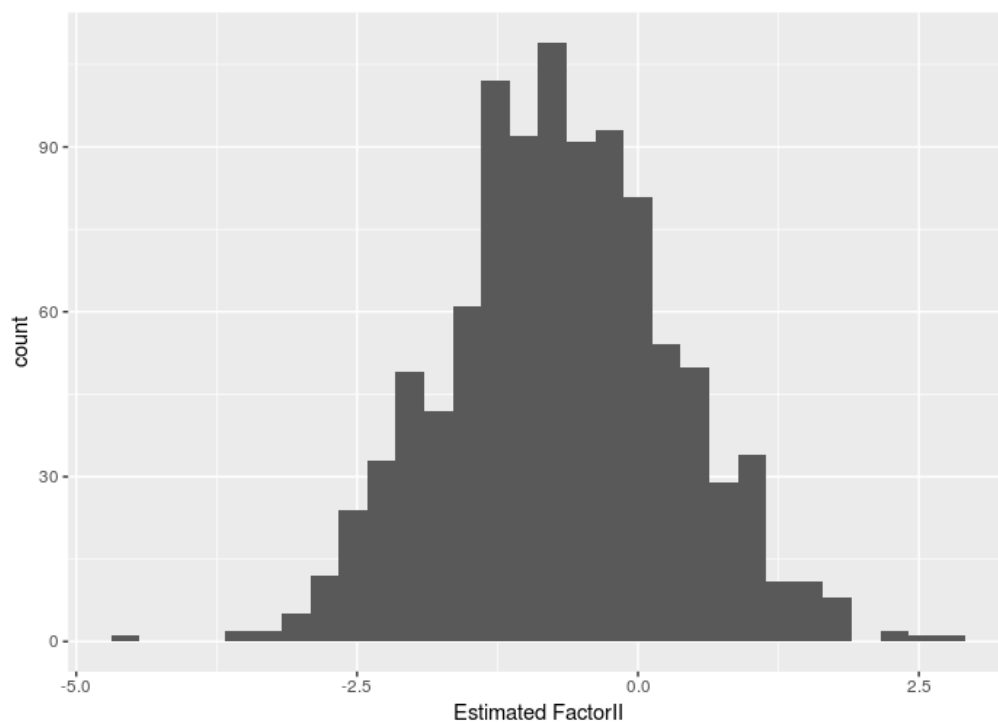


Figure 5: Posterior samples of estimated FactorII levels for a patient given their TEG values.

References

- Francis R Bach and Michael I Jordan. A probabilistic interpretation of canonical correlation analysis. 2005.
- Karim Brohi, Jasmin Singh, Mischa Heron, and Timothy Coats. Acute traumatic coagulopathy. *The Journal of trauma*, 54:1127–1130, 2003. ISSN 0952-7907. doi: 10.1097/ACO.0b013e3283509675.
- Marcus Brubaker, Mathieu Salzmann, and Raquel Urtasun. A family of mcmc methods on implicitly defined manifolds. In *Artificial Intelligence and Statistics*, pages 161–172, 2012.
- Simon Byrne and Mark Girolami. Geodesic monte carlo on embedded manifolds. *Scandinavian Journal of Statistics*, 40(4):825–845, 2013.
- Bob Carpenter, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 20, 2016.
- Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian data analysis*, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA, 2014.
- E Gonzalez, EE Moore, HB Moore, MP Chapman, CC Silliman, and A Banerjee. Trauma-induced coagulopathy: an institutions 35 year perspective on practice and research. *Scandinavian Journal of Surgery*, 103(2):89–103, 2014.
- John R Hess, Karim Brohi, Richard P Dutton, Carl J Hauser, John B Holcomb, Yoram Kluger, Kevin Mackway-Jones, Michael J Parr, Sandro B Rizoli, Tetsuo Yukioka, et al. The coagulopathy of trauma: a review of mechanisms. *Journal of Trauma and Acute Care Surgery*, 65(4):748–754, 2008.
- D Hoyert and Jiaquan Xu. National Vital Statistics Reports: Deaths: Preliminary Data for 2011. *National vital statistics reports*, 61(6), 2012. URL <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:National+Vital+Statistics+Reports+Deaths+:+Preliminary+Data+for+2011#4>.
- Plamen Koev and Alan Edelman. The efficient evaluation of the hypergeometric function of a matrix argument. *Mathematics of Computation*, 75(254):833–846, 2006.
- Matthew E Kutcher, Adam R Ferguson, and Mitchell J Cohen. A principal component analysis of coagulation after trauma. *The journal of trauma and acute care surgery*, 74(5):1223, 2013.
- Robb J Muirhead. *Aspects of multivariate statistical theory*, volume 197. John Wiley & Sons, 2009.
- Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- Pourzanjani, Jiang, Atzberger, and Petzold. “Fully Bayesian Inference of Dimensionality Reduction Models”, ”In preparation”.

Stef Van Buuren. *Flexible imputation of missing data*. CRC press, 2012.

Table of Patient Demographics

Table 2: Brief sub-cohort description including number of samples in each sub-cohort (N), percentage of the sub-cohort that is male, average and standard deviation of age, percent of patients with traumatic brain injury, and median blood units recieved in the sub-cohort.

Injury Mech.	N	% Male	Mean Age	Std. Age	% TBI	Med. Blood Units
GSW	85	90	27	8	40	5.0
MVC	52	90	27	7	75	0.5
SW	21	71	29	8	14	3.0
Assault	16	100	27	7	94	0.0

Table of Missing Values

Table 3: Number of missing values for each measurement type.

Measurement	Number of Missing Value (out of 174)
Proteins	
FactorII	49
FactorX	50
Protein C	44
D-Dimer	54
Fibrinogen	123
Platelets	8
TEG	
R	135
K	135
MA	133
Ly30	133

Use of Cauchy Priors for Sparsity

We also note that we place Cauchy priors over Λ and Γ which induce a sparsity property over latent dimensionality, forcing Λ and Γ to go to zero unless the signal in the data is sufficiently strong. Similarly we place Cauchy priors on the entries of the weight matrices W and B to induce sparsity in the matrix entries for interpretability purposes. These Cauchy priors serve a purpose that is akin to Laplace priors (or the L1 regularization in frequentist analysis), because just like the Laplace prior, the Cauchy prior places most of its mass near zero, with large mass in the tails of the distribution, allowing for appropriately large

values when statistical signal is strong enough. However, we prefer the Cauchy prior, as its continuity property makes more sense in the context of our analysis (see (Gelman et al., 2014) for a discussion on using Cauchy priors over Laplace priors). Finally, we mention that the hyper-parameter values of the Cauchy priors are chosen by leaving them as an unknown in our Stan model and conducting full Bayesian inference to sample from their posterior distribution.