Generalized Unrestricted Models (GUMs), a flexible and interpretable tool for behavioral and neural analysis

Alexandre Hyafil (alexandre.hyafil@gmail.com)
Universitat Pompeu Fabra, Calle Ramón Trias Fargas, 25
Barcelona 08018, Spain
Centre de Recerca Matemàtica, Campus UAB Edifici C
Bellaterra 08193, Spain

Vincent Adam (vincent.adam87@gmail.com)
Prowler.io, 72 Hills Rd,
Cambridge CB2 1LA, United Kingdom

Abstract:

Generalized Linear Model (GLMs) analysis is a popular tool in psychophysics and neuroscience for inferring the relative influence of various experimental factors onto choices, reaction times, neural activity and other observables. However, GLMs are intrinsically limited by their linearity assumption, and can lead to severe misattribution errors when (correlated) regressors contribute nonlinearly to the observed response. We show how this framework can be expanded to capture nonlinear functions. First, Generalized Additive Models (GAMs) allow to capture a nonlinear contribution for each regressor. A Gaussian Processes (GP) treatment of GAMs allows to recover the posterior distribution for each nonlinear mapping. Second, as neuroscience is often interested in the interaction of cognitive factors, we present a Bayesian treatment of Generalized Multilinear Models (GMMs) that capture multilinear interactions between different sets of regressors. GMMs can be applied e.g. when inferring the modulation of sensory processing by additional factors such as attention factors. Merging the frameworks of GAMs and GMMs yield Generalized Unrestricted Models (GUMs), a highly versatile and interpretable environment to capture cognitive determinants of behavior and neural activity. Crucially, these models can be efficiently estimated, even with limited dataset; Bayesian techniques can be applied to test which model is best supported by data.

Keywords: methods; regression analysis; modelling; psychophysics; Gaussian Processes

Limitations of GLMs

GLMs are represented by the generative model $E[y_k] = \sum_i w_i X_{ik}$. When regressors $(X_1, .. X_n)$ are correlated and the true underlying model is nonlinear, significant weights can be detected for non-contributing regressors. For example, we applied linear regression on spike count for a MT neuron during visualization of a random-dot kinetogram (Bair, Zohary, & Newsome, 2001), with motion coherence and accuracy of the monkey response at the end of sensory presentation as regressors. Surprisingly, this analysis concludes that

the neuron response activity depended on whether the animal would behave correctly (p<0.005, figure 1a). However, such false positive can be accounted for by the fact (figure 1b) that neural activity depends nonlinearly on motion coherence (full blue curve), and that accuracy (red curve) correlates with the residual (gray) from the linear trend (dotted blue line). Various enhancements of the GLM framework have been proposed to correct for these limitations, but are seldom used in cognitive psychology and neuroscience.
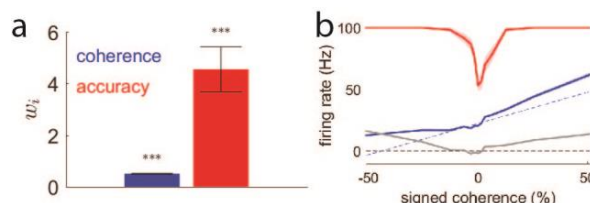

Figure 1: False positive in a GLM analysis induced by incorrect linearity hypothesis. a) Result of the GLM analysis with MT neuron spike count as predictor, and signed coherence and monkey accuracy on the same trial as regressors. b) Average neuron firing rate (blue curve) and monkey accuracy (red curve) as a function of signed coherence. Blue dotted line represents the linear trend, the gray line is the residual.

GAMs model each regressor as nonlinear mapping

GAMs allow to fit functions f_i in the form $E[y_k] = \sum_i f_i(X_{ik})$ (Wood, 2006). Traditional treatment of GAMs is limited however by the a priori definition of a set of basis functions for each regressor. GP framework sets prior $f_i \sim N(0, K_i)$, where K_i is a certain covariance function (e.g. Gaussian kernel) defining the expected variance and smoothness for f_i (formally the covariance for any finite set of function evaluations). Then an (approximate) posterior distribution is computed for the different mapping $p(f_i|y)$ (Adam, Hensman, & Sahani, 2016). In figure 2a we illustrate how a function $f_2(x) =$

$2cos(x)$ (red curve) is estimated (blue curve) by observing binary from model $p(y_k) = \sigma(X_{1k} + 2cosX_{2k})$ and fitting generative model $p(y_k) = \sigma(w_1X_{1k} + f_2(X_{2k}))$. Fitted smoothness parameter (e.g. by maximizing ELBO) naturally adapts to dataset; small dataset will lead to overestimated scale parameter, naturally retaining the most relevant trend in f_i (figure 2b). GP prior also improves weight estimation for ordinal regressors (i.e. inferring the influence of sensory samples in a sequence), compared to vanilla or L2-regularized GLM (figure 2c).
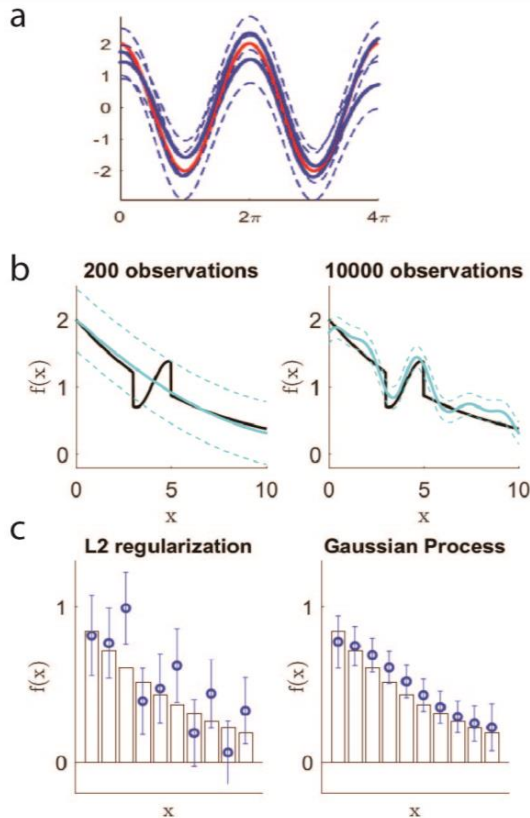

Figure 2: GAMs with GP framework allow to capture nonlinear mapping of regressor onto observables. a) Fit of a GAM model with 500 binary observations (blue curve: MAP function; dotted lines represent standard deviation of the posterior distribution) against true function (red). b) Impact of number of observations. With low number of observations (left), the fitted scale of the SE kernel remains high, essentially avoiding overfitting of the function (true function: black; estimated function: blue curve). With higher number of observations (right), the fitted scale goes down, allowing to capture finer details. c) Comparison of a model fits when binary observations (choices) derive from the weighted sum of weighted sum of evidence provided by sequential samples,

with exponential decay (referred to as *priming effect* in the literature). Applying GP prior (right) allows a better estimation of the weights than standard techniques such as unregularized GLM or L2-regularized GLM (i.e. ridge logistic regression, left panel).

GMMs model interactions between sets of regressors

GMMs use sets of regressors, e.g. $(R_1,..R_n)$ and $(S_1,..S_m)$: $E[y_k] = \sum_i u_i R_{ik} . \sum_j v_j S_{jk}$ (bilinear case) (Ahrens, Linden, & Sahani, 2008). While no longer convex, the fitting procedure for weights u_i and v_j (which involves alternatively updating one set of weights using Iterative Reweighting Least Square (Bishop, 2006) while leaving the others unchanged) generally gives consistent results. Compared to the full interaction model $E[y_k] = \sum_{i,j} w_{ij} R_{ik} S_{jk}$, they offer more interpretable results and require much less data, as the number of parameters scale as *O(m+n)* and not *O(mn)*. We illustrate the method on choice data in a perceptual accumulation 2AFC task where human subjects add to judge whether the overall orientation of a sequence of Gabor patches was tilted leftwards or rightwards (Wyart, de Gardelle, Scholl, & Summerfield, 2012). Standard GLM for responses R_k is $p(R_k) = \sigma(w_0 + \sum_i w_i X_{ik})$, where $X_{ik} = cos(\theta_{ik} - \theta^{ref})$ and θ_{ik} is the orientation for the ith sample in the sequence. GMMs with 4 interacting set of regressors (sensory sample, previous trial outcome, block position and subject id) captured modulation of sensory processing by previous trial outcome (after-error correction, second panel) and block position (improvement during learning followed by fatigue-induced performance decay, third panel): $p(R_k) = \sigma(w(subj(k)) + \sum_i w_i X_{if} f_2(accu(k-1) f_3(block(k)) f_4(subj(k)))$. GMMs analysis reveals jointly: recency effect (last samples in sequence have larger influence, figure 3a), after-error control adaptation (higher sensitivity following errors, figure 3b); modulation of performance by block (figure 3c); and heterogeneity in subject performance (figure 3d).
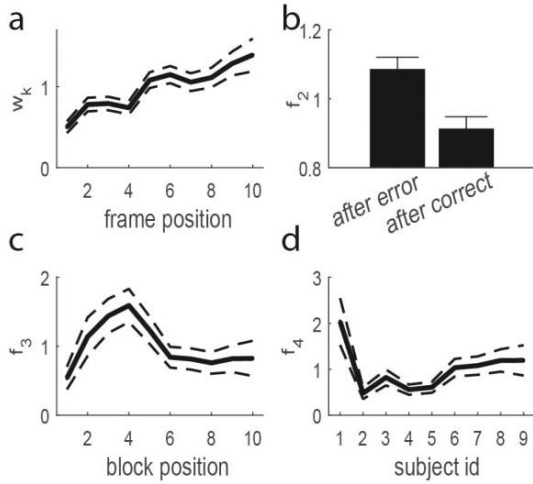
Figure 3: Fit of a GMM to participants choices in a visual accumulation task. a) estimated weight for visual frames (titled Gabors) as a function of position in the sequence w_i. b) modulation by previous trial outcome f_2. c) modulation by block index f_4. d) modulation by subject identity f_4 (9 subjects).

GUMs combine the framework of GAMs and GMMs

GUM general form is: $E[y_k] = \sum_c \prod_{d=1}^{m(c)} \sum_i f_{cdi}(X_{cdik})$, where each $f_{cdi}(X_{cdik})$ can be substituted by $w_{cdi}X_{cdik}$ if linearity with this regressor is assumed. This offers a particularly rich and flexible framework, where depending on the cognitive model one wants to test one can fit and compare models such as $E[y_k] = (f_1(X_{1k}) + w_2X_{2k})f_3(X_{3k}) + w_4X_{4k}$ or $E[y_k] = (f_1(X_{1k}) + f_2(X_{2k}))(f_3(X_{3k}) + f_4(X_{4k})) + \sum_{i=5}^{10} w_iX_{ik}$. Inference on function f_i and weights w_i can be run using a variety of Gaussian approximations, from Laplace approximation to variational inference and Expectation-Maximization (Rasmussen & Williams, 2006). Results on synthetic data using various observation types (binary, continuous, counts) showed that fitting is generally well behaved (not shown). We used GUMs in the accumulation task dataset to probe the sensory-to-perceptual mapping. Instead of assuming the optimal cosine mapping as in e), we estimated f_1 using GUM: $p(R_k) = \sigma(w(subj(k)) + \sum_i w_i f_1(\theta_{ik}) f_2(subj(k)))$. Fitted values show that the actual sensory-to-perceptual mapping is close to the optimal cosine mapping (figure 4b), as well as previously reported recency effect in sample weighting (figure 4a), and subject-dependent lateral bias $w(subj(k))$ (figure 4c)
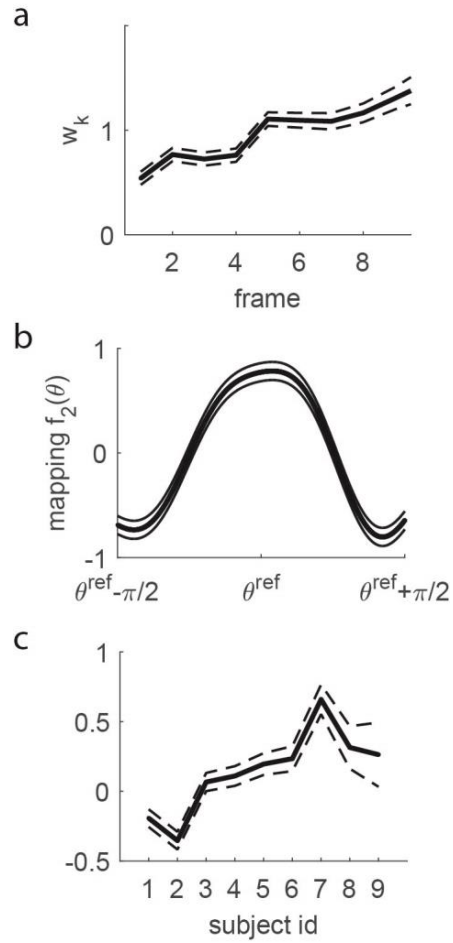

Figure 4: Fits of a GUM to participants choices in visual accumulation task (same data as figure 3). a. estimated weight for visual frames as a function of visual frames. b) nonlinear mapping from orientation to weight, closely following the normative cosine form. Maximal positive weight towards right responses is achieved when the orientation of the sample is aligned with that of the reference orientation for right responses (45 degree tilted rightwards), maximal negative weight for the opposite orientation. c) Modulation by subject identity f_2.

Conclusions

A Matlab toolbox will be available to allow for rapid, flexible usage of GUMs for cognitive psychology, neuroscience and other disciplines.

Acknowledgments

References

Adam, V., Hensman, J., & Sahani, M. (2016). Scalable Transformed Additive Signal Decomposition By Nonconjugate Gaussian Process Inference. In *Machine Learning for Signal Processing* (p. 1316).

Ahrens, M. B., Linden, J. F., & Sahani, M. (2008). Nonlinearities and Contextual Influences in Auditory Cortical Responses Modeled with Multilinear Spectrotemporal Methods. *Journal of Neuroscience*, *28*(8), 1929–1942. https://doi.org/10.1523/JNEUROSCI.3377-07.2008

Bair, W., Zohary, E., & Newsome, W. T. (2001). Correlated firing in macaque visual area MT: time scales and relationship to behavior. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *21*(5), 1676–1697. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11222658

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. *Vasa*. Retrieved from http://medcontent.metapress.com/index/A65RM03P4874243N.pdf

Wood, S. S. N. (2006). Generalized Additive Models: An Introduction with R. *Chapman & Hall, UK*, *63*(4), 410. https://doi.org/10.1111/j.1541-0420.2007.00905_3.x

Wyart, V., de Gardelle, V., Scholl, J., & Summerfield, C. (2012). Rhythmic Fluctuations in Evidence Accumulation during Decision Making in the Human Brain. *Neuron*, *76*(4), 847–858. https://doi.org/10.1016/j.neuron.2012.09.015