Application of Regularization Methods to Reconstruct Past Surface Temperatures

Evan Barrett
MATH 76.02: Computational Inverse Problems
August 2022

Introduction

Earth is currently undergoing a period of rapid warming, which has led to the melting of glaciers, sea level rise, migration of organisms, desertification of regions, and many other effects. As a result, it is important to understand how climate has changed in the past to have a better understanding of how the changes of today compare to those of the past. However, observed meteorological data going back more than two centuries is sparse, unreliable, or unavailable. In turn, scientists have developed several proxies for estimating past temperature, such as oxygen isotopes (^{18}O) and carbon dioxide concentrations in ice cores [1]. Oftentimes, these methods are used alongside other methods to check the results of one proxy against another. For this reason, the development of additional methods for reconstructing past global surface temperatures (GSTs) is useful. Borehole climatology is a new method that has been used recently to calculate past GSTs based on temperatures of the borehole at different depths [1]. By formulating this method as an inverse problem, one can attempt to uncover climatic history using regularization methods. The method has been applied for the climatic history of the past 2000 years [3], but longer histories have not been interpolated using these regularization methods. This project attempts to use these methods to calculate the climatic history of the past 100,000 years.

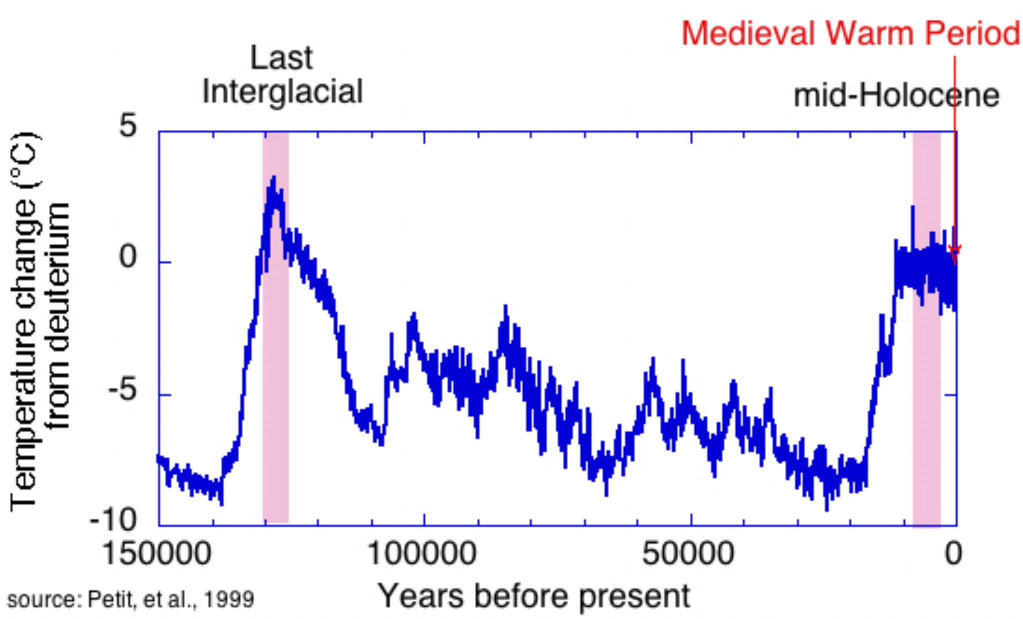

Figure 1: A climatic history of the past 150,000 years, determined from oxygen isotopes [4]

The Data

For this temperature reconstruction, we use the temperature of a Greenland Ice Core Project (GRIP) borehole from 1994 measured at depths up to 3000 meters alongside age-depth relationships measured from Greenland Ice Sheet Project (GISP) analyses [6]. Thus, we have the temperature and age of the ice at various depths z_j. We use 278 depth measurements ranging from 11.1 meters to 2781.1 meters below the surface, representing over 100,000 years of climatic history. The data was obtained from the National Oceanic and Atmospheric Administration's paleoclimate data search.

Methods

The problem of constructing past GSTs from borehole temperature data is constructed as an ill-posed inverse problem, which is then solved using a variety of regularization techniques.

The Inverse Problem

The diffusion of heat within frozen ground is controlled by a one-dimensional heat equation: $\kappa\frac{\partial^2 T}{\partial z^2} = \frac{\partial T}{\partial t}$, where κ is the thermal diffusivity of frozen ground (assumed to be constant in this case), z is depth, T is ground temperature, and t is the time before the present. The present temperature perturbation, given a past surface temperature $T_0(t)$, can be represented by the relationship

$$T(z) = \frac{z}{2\sqrt{\pi\kappa}} \int_0^\infty T_0(t) t^{-3/2} e^{-z^2/(4\kappa t)}$$

[3].

Our goal is to obtain $T_0(t)$ given some observed data $T(z_j)$ from a borehole. The above integral can be discretized and written as:

$$T(z_j) = \sum_{k=1}^{K} A_{jk} T_k$$

where $A_{jk} = erfc(\frac{z_j}{2\sqrt{\kappa t_{k-1}}}) - erfc(\frac{z_j}{2\sqrt{\kappa t_k}})$, where $erfc(x)$ is the complementary error function [3]. We thus obtain the matrix equation

$$Ax = b$$

where $x = (T_1, T_2, \ldots, T_K)^T$, which is a vector of the K GSTs temperatures at time k, and $b = (T(z_0), T(z_1), \ldots, T(z_N))^T$, which is a vector composed of the N observed borehole temperatures at a depth of z_j [3].

Using the data obtained from the GRIP borehole data set, we construct the matrix A and obtain a condition number of $2.0466 * 10^{27}$, demonstrating the ill-posedness of the problem. As a result of the ill-posedness, the problem cannot be solved by simply multiplying both sides by A^{-1}. We treat potential errors in the borehole temperature data as noise and apply a variety of regularization techniques to reduce the effects of noise. All of the regularization parameters and the resulting solutions are computed within MATLAB using the `Regularization Tools` and `cvx` packages.

Least Squares Method

The simplest attempt at finding the GSTs involves solving a least squares, which is the same as minimizing $\|Ax_\lambda - b\|_2^2$ [2]. To do so, we utilize the slash function in MATLAB and calculate x as `x=A\b`. The resulting reconstruction is shown in Figure 2. As seen in the figure below, the least squares reconstructing does a decent job of capturing some of the intricacies of the climate record compared to the raw borehole temperature data. The least squares method is thus a valuable method to try. If A was less ill-conditioned, then the results of the least squares method may be closer to those of the other proxies for paleoclimate temperatures.

Choosing a Regularization Parameter

For the following regularization methods, we need to choose a regularization parameter λ that balances $\|Ax - b\|$, which measures how well the solution x predicts the noisy b, and $\|x\|$, which measures the regularity of the solution [2]. To choose a regularization parameter, we employ a variety of methods.

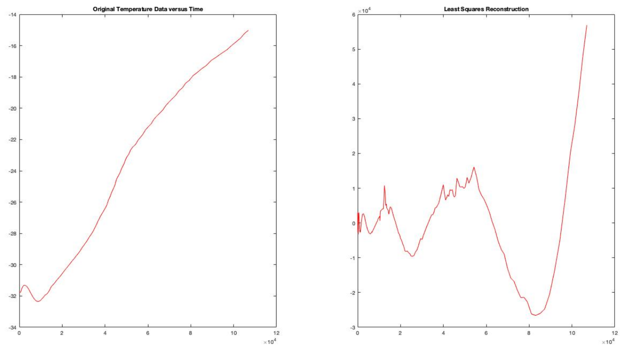

Figure 2: The reconstructed GST data versus time (right) compared to the original temperature data (left).

We first use the L-curve method, which uses a log-log plot of the solution norm $\|x_\lambda\|$ versus the residual norm $\|Ax_\lambda - b\|$. From this curve, we use the point of greatest curvature to determine our regularization parameter. The L-curve for the problem is shown below in Figure 3. As seen in the figure, the obtained regularization parameter from this method is $\lambda = 0.1326$.

We then use the discrepancy principle to find another possible regularization parameter. The discrepancy principle finds a parameter such that $\|Ax_\lambda - b\|$ equals the discrepancy in the data, which is based on the underlying noise of the data set. We thus have to make some sort of prediction about the magnitude of noise within the system. We choose a constraint $\delta = 0.01$ based on knowledge of the potential error within the data. Using the `discrep` function, we calculated a regularization parameter of $\lambda = 5.0441 * 10^{-19}$.

The next method for calculating the regularization parameter is generalized cross validation (GCV) analysis, which involves trying to calculate a parameter λ such that x_λ predicts the de-noised version of b, b_{exact}. The regularization parameter is one that minimizes the expression $\frac{\|Ax_\lambda - b\|_2^2}{(m - \sum_{i=1}^{n} \varphi_i^{[\lambda]})^2}$, where $\varphi_i^{[\lambda]}$ is a filter factor and m is the number of observations [2]. The plotted GCV analysis is shown in the center of Figure 3. The calculated regularization parameter from this method was $\lambda = 1.1760 * 10^{-15}$.

Our last method for calculating the regularization parameter is normalized cumulative periodogram (NCP) analysis, which involves determining when the residual becomes dominated by the noise rather than the data itself. To do so, the NCPs for a variety of regularization parameters are plotted alongside each other, and the one closest to a straight line is chosen as the optimal regularization parameter. A NCP that is a straight line implies that the residual vector resembles white noise, and thus we choose that regularization parameter to balance the residual and the regularity of the solution [2]. The plotted NCPs are shown in Figure 3 on the right. As seen in the figure, the calculated optimal regularization parameter is $\lambda = 1.1760 * 10^{-15}$.

We thus have three different regularization parameters to test: $\lambda = 1.1760 * 10^{-15}$, $\lambda = 5.0441 * 10^{-19}$, and $\lambda = 0.1326$. We also use the regularization parameter $\lambda = 0.001$ as obtained by Liu et al., (2019) in a similar analysis.

Tikhonov Regularization

With our regularization parameters, we reconstruct past GSTs using the Tikhonov Regularization method with the borehole temperature data. This involves finding the value of x_λ that minimizes $\|Ax_\lambda - b\|_2^2 + \lambda^2 \|x_\lambda\|_2^2$ for each of our regularization parameters λ.

As seen in Figure 4 below, the regularization parameter that best matches the established temperature record appears to be $\lambda = 0.001$. When the

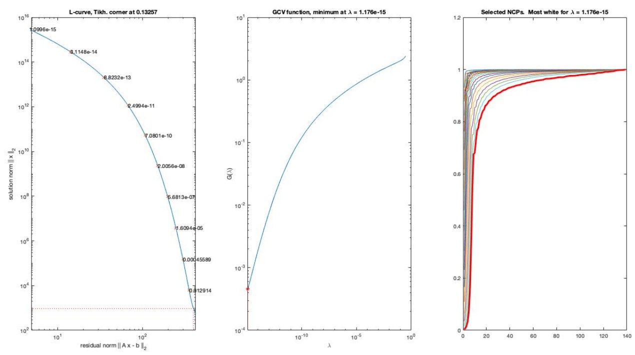

Figure 3:
Three methods for choosing the regularization parameter λ based on the L-curve (left), GCV Analysis (center), and NCP Analysis (right).

regularization parameter is too low, the noise overtakes the data and prevents the temperature trends from showing up in the reconstructions. However, if the regularization parameter is too high then the borehole temperature data is not filtered enough and thus does not reveal the trends in the GSTs as well. Therefore, a good regularization parameter is important for computing the GSTs from the borehole data. The fact that the analyses conducted to determine an ideal regularization parameter did not yield one of 0.001 may be due to the ill-posedness of the system. Liu et al., (2019) cross-referenced their borehole temperature data with modern data, allowing them to get a better correlation between the two and thus get a stronger regularization parameter. Future studies for multi-millennia temperature reconstructions could attempt to use modern meteorological data, if available, to constrain the regularization parameter.

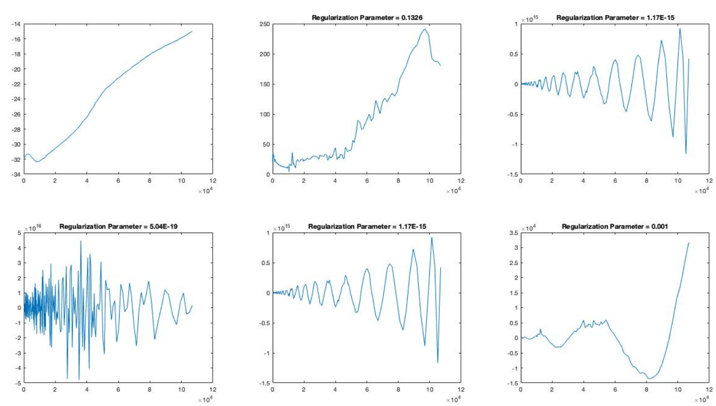

Figure 4:
GST reconstructions using Tikhonov regularization with four different regularization parameters, in comparison to the original borehole temperature data (top left.)

L1 Regularization

L1 regularization techniques were also used to attempt to reconstruct the past GSTs. L1 regularization is generally used when the underlying system is composed of sparse signals, which may allow for this method to work based on the fact that climate can be controlled by seemingly sparse, rapid changes [2]. With L1 regularization, we seek to find the x_λ that minimizes $\|Ax_\lambda - b\|_2^2 + \lambda\|x_\lambda\|_1$. As seen in Figure 5, the L1 regularization does not do a good job of capturing the complete GST history due to its ability to only capture sparse signals. However, it is worth noting that two out of four of the regularization parameters led to a reconstruction at which there was a small, negative sparse signal at around 12.5k years before present. The sparse signal at this time could be correlated to the abrupt Younger Dryas cooling period, which was a period of cooling during the time of deglaciation in the Quaternary period. Thus, L1 regularization could potentially be used as a tool for identifying these sparse, abrupt changes in climate that scientists are still actively working to understand.

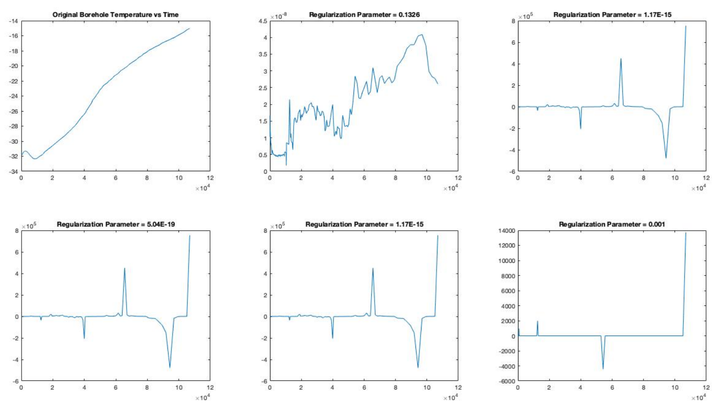

Figure 5: GST reconstructions using L1 regularization with four different regularization parameters, in comparison to the original borehole temperature data (top left.)

Total Variation Regularization

The final regularization method that was used to interpolate GSTs from the borehole temperature was total variation (TV) regularization. TV regularization works well when the expected underlying solution is discontinuous. The method looks at how each data points differs from its neighbor and aims to minimize the difference between two adjacent data points. If we are assuming that the underlying solution is piecewise continuous, then we would expect the total variation to remain roughly zero for most data points and should only be nonzero when a jump discontinuity occurs. We thus aim to find the solution x_λ that minimizes $\|Ax_\lambda - b\|_2^2 + \lambda TV(x)$. We find the regularized solution for each regularization parameter using the `cvx` package in MATLAB. As seen in the resulting graphs in Figure 6, the solutions calculated from TV regularization methods generally do not show the overall climate history. With low regularization parameters, the regularized solution adds jumps in the data set that are not seen in other records of the climate. The higher regularization patterns show less jumps in the regularized solution; however, the jumps in these solutions also cannot be seen in established climate records. For example, the solution x_λ where $\lambda = 0.001$ shows a rapid increase in temperature at about 20,000 years before present, which is not consistent with the climate record and evidence of glaciation during that time. Therefore, the TV regularization method is not reliable for climate reconstruction purposes. Refining of the method and the kernel matrix could possibly allow TV regularization to highlight overall trends in the climate record, such as periods of glaciation versus periods of deglaciation; however, the current method does not show those trends in the calculated GSTs.

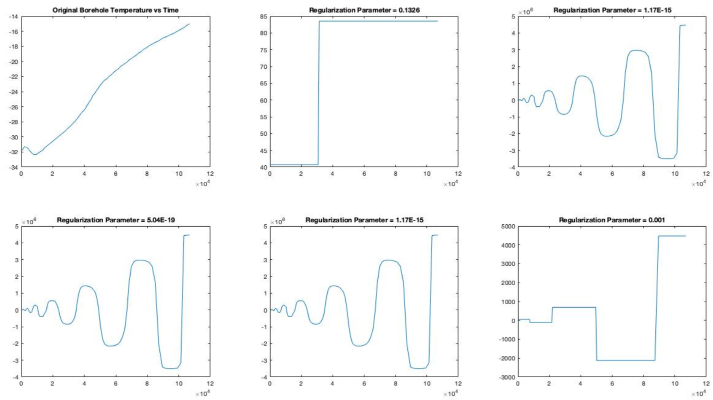

Figure 6: GST reconstructions using TV regularization with four different regularization parameters, in comparison to the original borehole temperature data (top left).

Truncated SVD Method

While the Truncated SVD (TSVD) method was ultimately not used, it is worth discussing. The TSVD method works by discarding the noise-dominated SVD coefficients [2]. While this method can reduce the amount of noise and decrease the ill-posedness of the problem, it also has the potential to remove critical information through the truncation. We conducted an analysis of the singular values of the matrix A through a Picard plot, shown below in Figure 7. As seen in the plot, the singular values rapidly decrease and nearly 80% of the singular values would need to be truncated to significantly reduce the condition number of the problem; however, the truncation would lead to the loss of valuable information regarding the climate trends. Therefore, we decided not to move forward with the TSVD method.

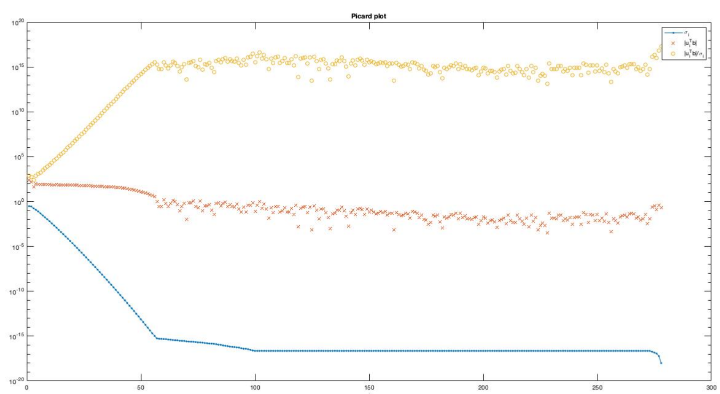

Figure 7: The Picard plot for the matrix A

Conclusions

This project aimed to develop a better understanding of how various regularization techniques could be used to interpolate past GSTs from present-day borehole temperatures. Though temperatures from the past 600 years had been previously calculated using these methods [3], there were no published attempts to reconstruct GSTs from the past 100,000 years using Tikhonov regularization or other regularization methods. The regularized solutions did not completely match the paleoclimate temperature trends that have been constructed using other proxies, most likely due to the high condition number of the kernel matrix A and the general ill-posedness of the problem. The regularized solutions have temperature values that are high in magnitude, most likely due to the ill-posedness. However, the Tikhonov and Least Squares methods seem to do a decent job of capturing overall trends, especially within the first 20,000 years of climate history before present. Furthermore, L1 regularization may have the potential to highlight drastic, sparse changes in climate that have a large global or hemispheric impact.

Furthermore, this project highlights how important the choice of the regularization parameter is. Through the use of a variety of methods for choosing the regularization parameter, we were able to plot many solutions side-by-side and compare them to the established paleoclimate temperature record to evaluate the effectiveness of the parameter. Parameters that are too low result in the excess of noise in the regularized solution, while parameters that are too high remove critical data from the problem that erases potential trends in the data. It is important to consider a wide range of regularization parameters and methods for choosing the parameter. Future studies could incorporate observed meteorological data to constrain the regularization parameter.

The results of this study could be improved upon by applying a Bayesian approach to inverse problems. In doing so, one could implement a prior probability distribution based on underlying knowledge of climate forcings to attempt to reconstruct the GST record from the borehole temperature. Applications of uncertainty quantification could also be used to reduce the amount of uncertainities within the borehole temperature data and thus make it more reliable for constructing paleoclimate records. Additionally, the development of a matrix A with a lower condition number than the matrix used in this project would help reduce the ill-posedness of the problem. Thus, it would be helpful to develop a better understanding of the relationship between past GSTs and current borehole temperatures and the factors involved in that relationship. If those factors could be reflected in the kernel matrix, then reconstructions may be more accurate.

The field of paleoclimate is increasingly important as the present climate continues to change, and increasing the amount of tools that are available for understanding past GSTs helps scientists and the general public understand how climate changed in the past and how those past changes may reflect present day changes. Thus, it is important to work towards refining this method of temperature reconstruction to increasing our understanding of the world and its past to protect its future.

References

[1] Bodri, L., & Čermák, V. (2007). *Borehole climatology: a new method on how to reconstruct climate* (1st ed.). Elsevier.

[2] Hansen, P.C. (2010). *Discrete inverse problems*. Society for Industrial and Applied Mathematics SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104.

[3] Liu, Zhang, T., Clow, G. D., & Jafarov, E. (2021). Application of Tikhonov regularization to reconstruct past climate record from borehole temperature. *Inverse Problems in Science and Engineering*, 29(13), 3167–3189. https://doi.org/10.1080/17415977.2021.1975700

[4] Petit, Jouzel, J., et al. (1999). *Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica*. Nature (London), 399(6735), 429–436. https://doi.org/10.1038/20859

[5] Shen, Beck, A. E., & Anonymous. (1990). Least squares inversion of borehole temperature measurements in functional space. Eos (Washington, D.C.), 71(43), 1603–.

[6] Johnsen, S., et al. (1995). Greenland palaeotemperatures derived from GRIP bore hole temperature and ice core isotope profiles. *Tellus*, 47(5), 624-629. doi: 10.1034/j.1600-0889.47.issue5.9.x