

Homework 8: Anomaly Detection

Due November 8, 2018 at 2:10 PM

This homework will explore PCA models for anomaly detection and diagnosis.

hwk8data (available on Canvas) includes three matrices - *data*, *test1*, and *test2* – which include 21 variables each. The matrices contain data collected from a Slurry Red Ceramic Melter (SRCM): the first twenty variables are temperature readings from thermowells throughout the melter and the 21st variable is the glass level. The temperature readings are in units of *°C*, while the glass level reading is in *inches*.

Here, we will use principal component analysis to determine if the data included in *test1* and *test2* matches with the data in *data*.

1. Briefly describe PCA and how it is used for anomaly detection. Describe the use of Q- and T²-statistics for detecting anomalies. Give the equations for calculating the Q- and T²-statistics.
2. Develop a PCA model of the normal operating data contained in *data*. Determine on your own how many PCs should be included in the model. Describe how you made this decision and why it makes sense! Using this model and the fault-free data in *data*, determine thresholds for the Q- and T²-statistics. Explain how you selected thresholds.
3. Project the data in *test1* and *test2* into the PC space that you found in step 2. Evaluate the T²- and Q-statistics to identify which, if any, observations are anomalous.
4. For data observations that are flagged as anomalies, look at the contributions to the T²- or Q-statistics (whichever is appropriate) to determine which sensor readings might be causing the anomalies. Visually inspect the raw signals for these to see if the faults are apparent.

For this homework, prepare a written report in IEEE format. Include any plots and tables that will support your findings. Make sure you correctly label your figures and tables and refer to them in the text. Include an appropriate citation for the data, both in the text and in the list of references. Your report should include **at a minimum** an abstract, introduction, methodology, results (and discussion!), conclusions, and references. Note that the methodology section of this report (and every report!) should describe the algorithm that you're using – not the implementation in MATLAB. Include all your code in an appendix (single column) at the end of the report. Convert your report to .pdf before submitting it through Canvas.