**1. Background and Introduction**

In contemporary society, power is needed everywhere in our daily life. Although not familiar to most of us, to create power, the generator needs momentum from the engine that produces kinetic energy from some sources. One type of combustion engine, the gas turbine, is a crucial component of the power generating process across all industries, such as aircrafts, trains, and electricity plants. Connected to the generator, the gas turbine drives the generator to function by pressurized gas: the air is drawn to the turbine from its suction and heated by fuel source combustion; then, as the heated air expands, the motion is made by the turbine and is connected to the generator to produce electricity.

Transforming pressure energy to kinetic energy, the turbine blades are the key element of a gas turbine. As it extracts energy from the high temperature and high pressure gas, it is also the limiting part of a gas turbine, as it needs to withstand the severe environment. To protect the turbine blades, an internal cooling system is incorporated in the turbine, extracting cooling air from the compressor and passing through the airfoils to cool the blades.

While blade deformation to some extent is acceptable, an extreme deformation can largely reduce the length of the engine life cycle, and the life of the blades themselves will decrease by half only by a 30 degree celsius off in the blade temperature prediction. As a result, the design of the turbine blades requires careful consideration on the choice of materials and the cooling schedule. Based on these two aspects, six components need to be examined: Young’s modulus (which quantified how the material can stretch under stress), Poisson’s rate (which measures how the material tends to expand perpendicular to the direction of compression), Coefficient of thermal expansion (abbreviated as CTE, which measures the relationship between material expansion and temperature), thermal conductivity, internal cooling air temperature, and air pressure load on suction.

With all the considerations on metal materials and cooling schedule, we aim to find an optimal design for the turbine blades to minimize blade stress and deformation, measured by displacement. To mathematically formulate our purpose, there are two goals we need to achieve: firstly, we want to improve the predictive accuracy of the blade stress and displacement. In other words, modeling our data with a Gaussian Process, we want to find the parameters that give us the most similar outputs to those from the pressure and displacement simulator, provided by our aerospace collaborators. Secondly, we want to optimize the acquisition of stress given the constraint on displacement of 1.3 × 10-3.

To achieve these two goals, we first select the best space-filling method for sampling. Using this method, we generate the initial model to fit our data. Then we update the Gaussian Process based on appropriate acquisition criteria on previously mentioned six features to obtain a better approximation of stress and displacement under the displacement constraint. Finally, we approach our second goal by optimizing the black-box simulator function with the real-valued constraint.

Through the process of this project, there are some limitations of time, computational, and budgets we need to consider. With limited sources, the project is limited to a total of 150 computational experimental runs. In addition, since the simulation process incorporates the simulator MatLab function from our collaborators, only some devices in our group can integrate the function to the rest of the code performed in RStudio.

**2. Method**

**2.1 First Goal - Obtain Evaluation Points**

To achieve our first goal, we need to select a set of *N* evaluation points within the constraints of the six features and the constraint on the displacement. We could use a model-based design to select all points, which allows us to control what design points we want by specifying hyperparameters of the Gaussian Process model. However, as a model-based design assumes perfect specification of the parameters to incorporate our prior beliefs, it is at risk of model misspecification and time consuming. For a safer, more robust, and faster approach, instead of selecting a batch of points at once, we first choose a space filling method to obtain a set of initial *n* evaluation points and then use a model-based sequential design to select the next points at a time that better approximates the black-box simulator function, until we reach the number of *N* points in total.

**2.1.1 Initial Design - Space Filling**

**2.1.2 Sequential Design**

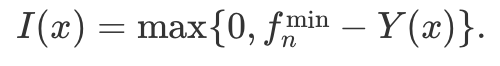
After obtaining the initial *n* points, we evaluate these points by the simulator function to obtain a complete data set with explanatory variables and corresponding responses. Then, we fit the *n* evaluation points to a Gaussian Process model and solve the objective function based on a chosen criterion to obtain the *n+1* point. With this new point, we observe a new response by the simulator to obtain a new pair of explanatory variables and response. We then add the pair to the whole data set, and repeat the process from fitting a new Gaussian Process and solving the objective function. We stop selecting new points until we have a total set of *N* points.

In the sequential design, we mainly have two criteria to choose new points: the Active Learning Mackay (ALM) and the Active Learning Cohn (ALC). ALM criterion aims to find the next point to maximize predictive variance based on the current dataset, that can also be considered as maximization of information gain given the assumed GP model, while ALC criterion applies a more generic metric which minimizes the integrated deduced variance. In other words, ALC criterion maximizes reduction in predictive uncertainty. Moreover, practically ALC gets to lower out of sample RMSE faster than ALM (shown in 6.2.2 Figure 6.16). Thus, we choose ALC criteria instead of ALM.

**2.2 Second Goal - Optimization**

* **Criterion: Expected Improvement**

On the other hand, Gaussian Process is useful in optimizing blackbox objective function, named surrogate-assisted optimization. One criteria in choosing next point is EY, which aims to minimize the fitted predictive mean surface in GP. Although it performs much better than the direct optimization, when we have random initializations, this criteria will not always be the global optimization tool, but sometimes stick with local optima (shown 7.1.1 Figure 7.4), since it hardly explores places that cannot be easily reached from current best value and highly depend on the problem itself.

A more advanced criteria, Expected Improvement(EI), also called efficient global optimization(EGO), takes not only predictive mean but also predictive uncertainty into account. It introduces the measure of improvement and aims to maximize the expected potential improvement based on current dataset. This criteria explores intervals of input space with great potential of finding minima but not stick to current minimum point, which is much more efficient and balance between exploration and exploitation better than EY. Practically, EI achieves lower objective value quicker than EY (shown in 7.2.2 Figure 7.11), thus we choose EI as our criteria in optimization.

Reference:

<https://www.energy.gov/fecm/how-gas-turbine-power-plants-work>

<https://studentenergy.org/conversion/gas-turbine/?gclid=CjwKCAjw9e6SBhB2EiwA5myr9kWNx5UoWzfanMTnDR37E8ru-N76QTHkdA5Acukb1MBZ43dCp9cPIBoCRAQQAvD_BwE>

<https://en.wikipedia.org/wiki/Gas_turbine#:~:text=Gas%20turbines%20are%20used%20to,%2C%20gas%20compressors%2C%20and%20tanks>.