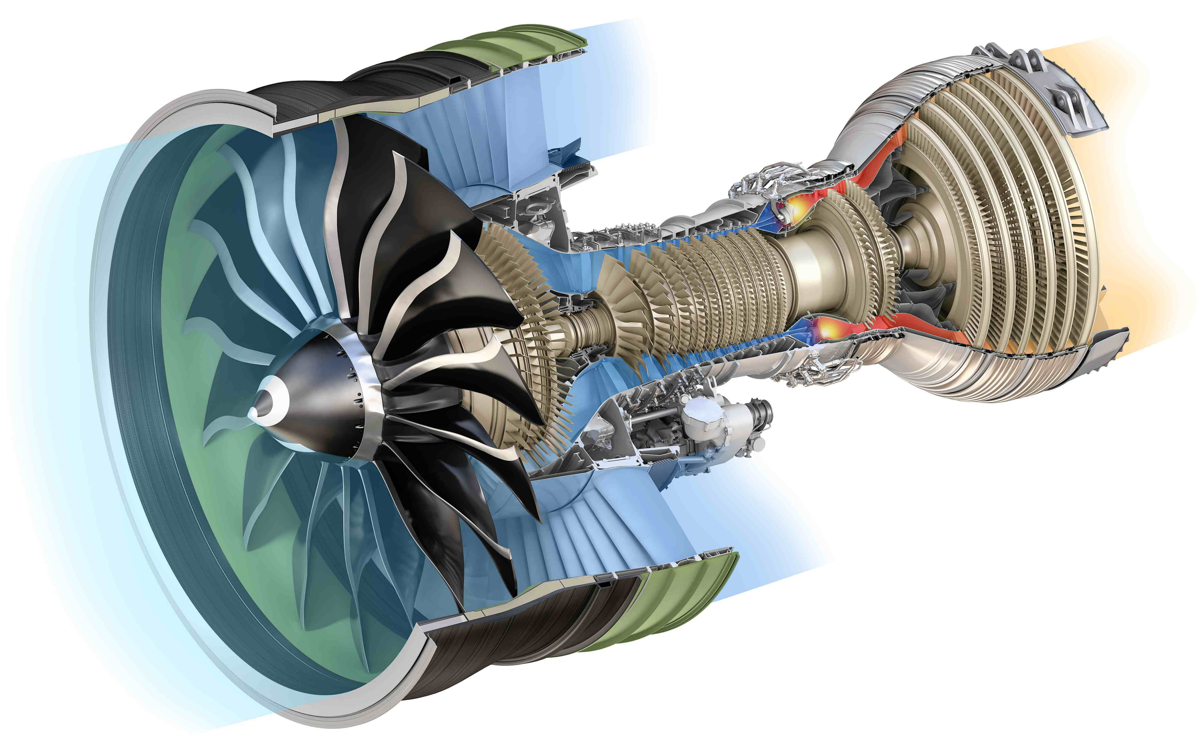
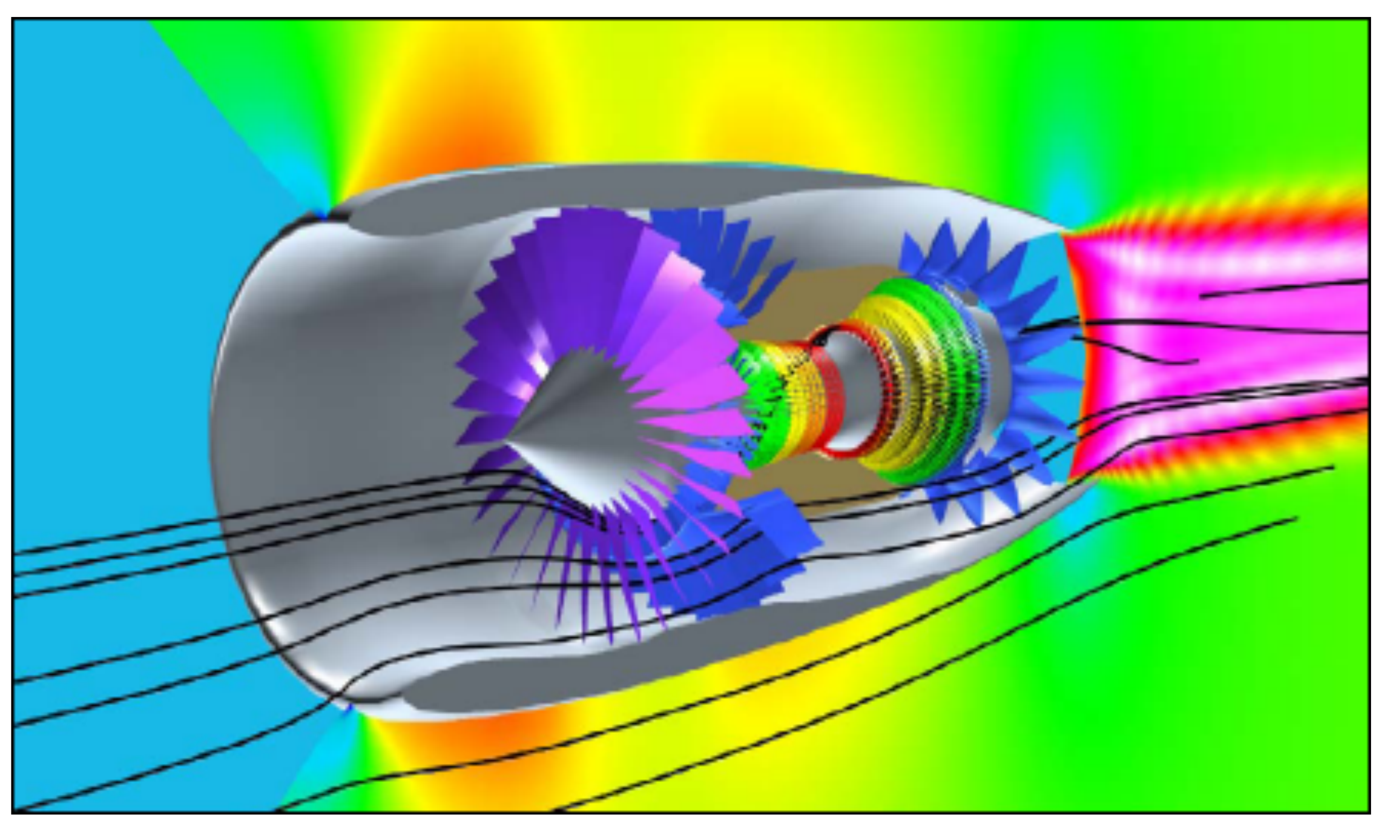
Probabilistic Adjoint Sensitivity Analysis for Fast Calibration of PDE Models

Jon Cockayne and Andrew Duncan

Parameter Estimation in Engineering

It is common to build **digital twins** of complex engineering systems. These are computer models often formed of complex sets of differential equations. Generally the models have **parameters** that need to be tuned to optimise some objective - perhaps to maximise the performance of a system in design, or to fit collected data.

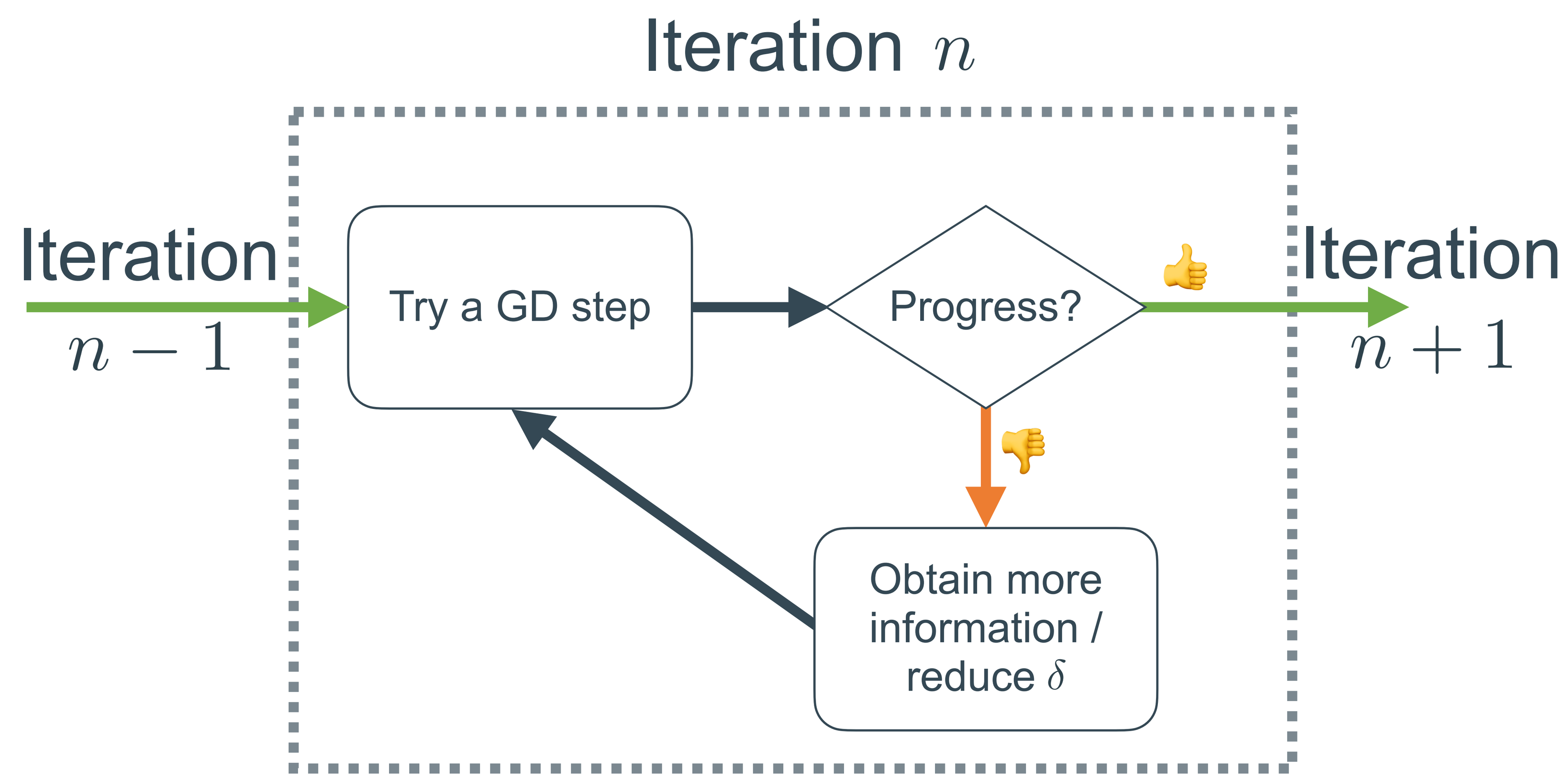




Gradient descent is among the most basic techniques for optimisation. In each iteration one must evaluate both the **objective function** and its **gradient** in order to take a step towards the optimum. Computing the gradient is a **local sensitivity analysis** problem (Bonnans and Shapiro, 2013) that can be extremely expensive!

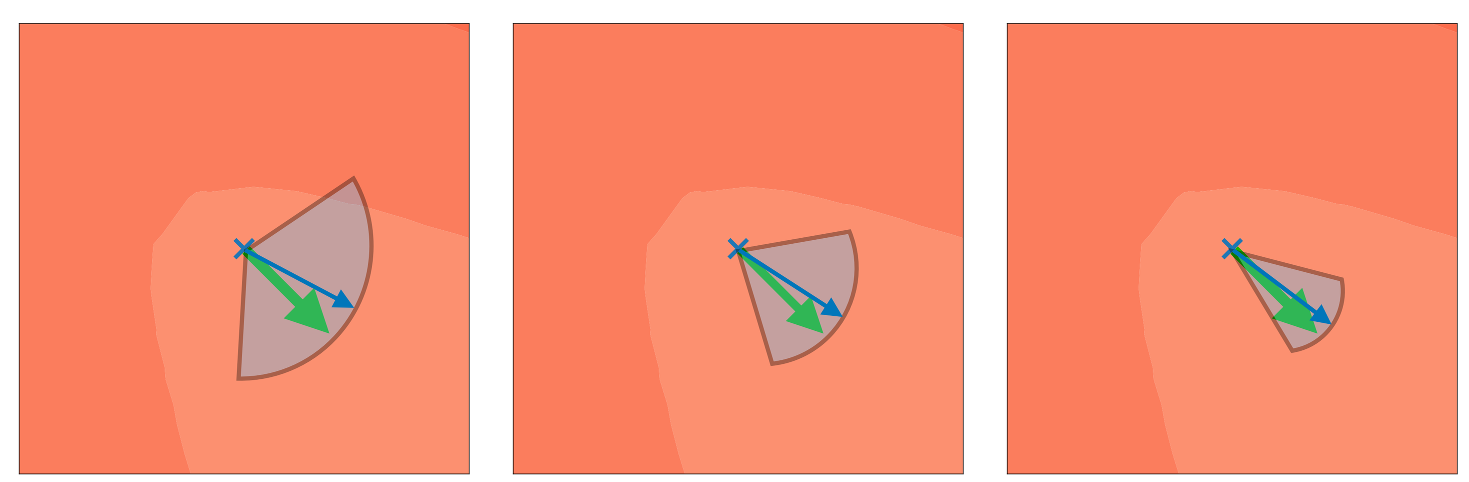
Probabilistic Gradient Descent

In our new approach (Cockayne and Duncan, 2020), we build a **global Gaussian process model** for the gradient field over parameter space. This can be incorporated into gradient descent to **reduce the cost** of computing gradients.



At each new parameter location the true gradient is approximated by the probabilistic gradient. The **mean** gives a best guess of the gradient and the **covariance** provides an error estimate. A demand on the size of the covariance at each iteration forms an input parameter, δ, that is dynamically varied during execution to ensure convergence.

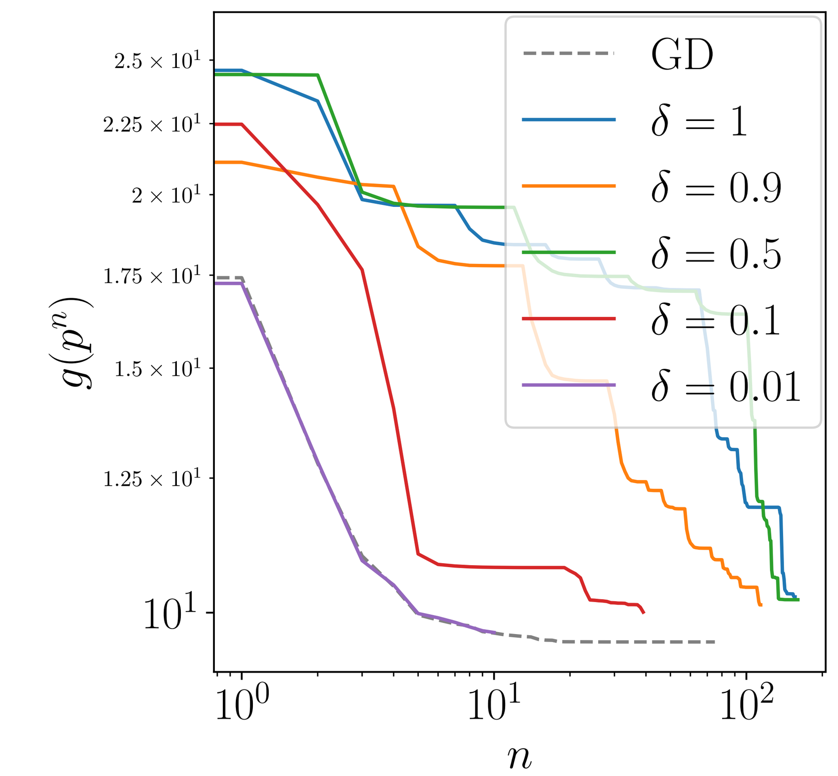
As the amount of information increases the cost of computing the probabilistic gradient approaches that of computing it exactly - but we can often compute an estimate that is "good enough" at a **lower cost**.



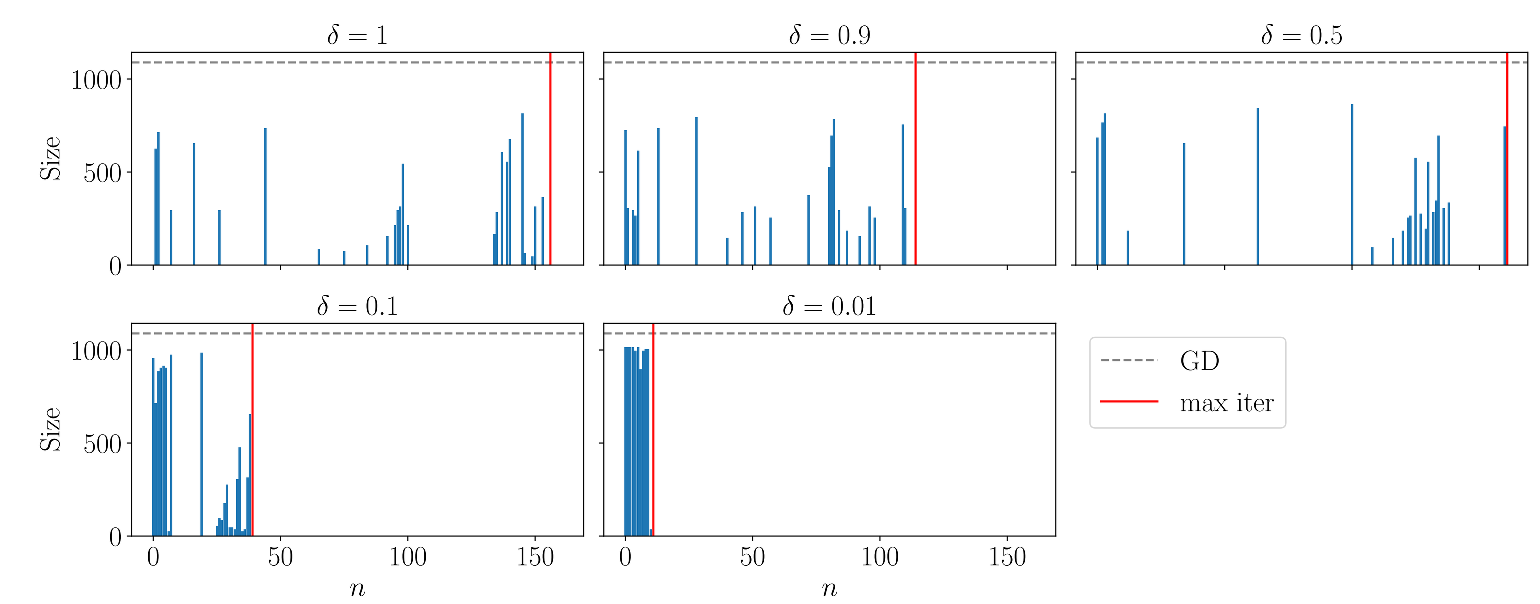
Eventually the gradient estimate is sufficiently accurate for a gradient descent step to be performed. However - because the gradient model is **global across parameter space** information from iteration i is transferred to iteration $i+1$. This has the effect of **further lowering the cost**.

Results

We applied probabilistic gradient descent to optimise an objective function $g(p)$ for a simple test PDE.



Clearly, larger δ makes the algorithm take more iterations to converge.



However - less information is collected for larger δ so that execution time - in terms of wall time - may be reduced.

References

Cockayne and Duncan (2020). "Probabilistic Gradients for Fast Calibration of Differential Equation Models." Preprint. https://arxiv.org/abs/2009.04239

Bonnans and Shapiro (2013). "Perturbation analysis of optimisation problems." Springer Science & Business Media.