CASE STUDY

Reverse-mode AD = **Faster Risk Management, Better Pricing, and Cheaper Compute**

How can banks best manage CVA risk to protect revenue, remain agile under pressure, and compute accurate regulatory capital charges? How do we place customers at the centre of our derivatives business? These are the tough questions one investment bank sought to answer with a highly innovative new XVA engine, leading them to choose NAG as a key partner.

These questions raise a few key points:

- Fair, accurate and keener pricing includes incremental impacts on hedges and valuation adjustments.
- Robust risk management requires rich sensitivity information and fast "What-if" capabilities.
- The imminent FRTB-CVA capital charges penalise banks who cannot hedge CVA across the multitude of risk factors stipulated by the regulators.

Fast and accurate sensitivity data is a common theme, with banks turning to automatic differentiation (AD) and reverse-mode AD (also known as backpropagation or adjoint AD) to produce fast sensitivities at reduced cost.

NAG's customers typically realise a 20x -300x reduction in compute costs when switching from bumping to automatic differentiation.

Our client's new XVA engine won a Technology Innovation Award and is a leap forward in their risk and pricing capabilities. Selecting NAG as a partner, they built the risk system leveraging NAG's products and expertise in AD, task orchestration, GPU accelerators and core analytics. The XVA system is designed to enable the best possible service to customers, protect revenue and reduce risks to the bank and its stakeholders.

The new system takes a holistic approach to pricing and risk management and has profoundly impacted the bank's trading activities. It is so fast that incremental XVA, including sensitivities, can be calculated on a per-deal basis, allowing traders to see the impact of a trade on valuation adjustments and hedges, and to use this information in pricing. The result is fairer, better pricing for customers and a continuous view of how the desk's risk is evolving. Full CVA revaluation with all sensitivities can be done in 30 minutes or less, giving the CVA desk rich insights into the bank's credit exposure.

On-demand "What-if" scenarios allow the CVA desk to evaluate the impact of likely events even when markets are moving rapidly, giving advance warning of potential issues.

The Need for Speed

Our client built a GPU-accelerated risk engine to perform XVA calculations in minutes, adding very fast sensitivities through backpropagation. This provides the XVA sensitivities dramatically faster than bumping. For CPU code (calibration, curve building, etc) the team selected the NAG® dco Suite, the industry-leading suite of AD tools and libraries engineered to make robust backpropagation simple.

Despite this, they still faced a serious challenge: CPU backpropagation without the corresponding GPU backpropagation would severely limit performance, and at the time there was no AD tool capable of handling general purpose CUDA code. In response to this need, NAG's AD and CUDA teams collaborated to build a new breed of AD tool to handle a modern XVA code.

Built on "meta adjoint programming" the new tool, NAG® dco/map, adopts a novel approach to the reverse-mode AD problem and works within the resource constraints of an accelerator. The combination of NAG® dco on the host and NAG® dco/map running on NVIDIA GPUs on Microsoft Azure allowed our client to do XVA valuation with hundreds of sensitivities in roughly 30 minutes rather than several hours (or days). Incremental XVA can be done in seconds, making it fast enough for traders to sharpen their pricing with this credit and sensitivity information.

Backpropagation (also known as reverse-mode AD or adjoint AD) "bumps the output" and propagates the changes backwards through the code to observe the effect on the inputs, except everything is computed analytically and results are accurate to machine precision.

Fast and Flexible

More banks have embedded NAG® dco/c++ into their quant libraries than any other AD tool. NAG® dco/c++ computes the sensitivities of arbitrary C++ codes using forward or reverse mode AD, or any combination, and can compute higher order derivatives as well. The sensitivities are more stable and accurate than finite differences, and for codes with many inputs, they can be dramatically faster: customers have reported speedups between 20x - 300x.

Flexibility is almost always the difference between success or failure when applying any AD tool to critical applications. Naïve backpropagation will consume vast amounts of memory. Memory use can be effectively controlled by using various checkpointing and Jacobian preaccumulation techniques, but these require access to the internals of the AD tool, for which NAG® dco/c++ has a powerful and intuitive API. Users can constrain memory use arbitrarily, integrate handwritten adjoints, and perform mixed forward-reverse mode AD to fully exploit the structure in their algorithms. NAG® dco/c++ also has vector types for AVX vectorization and for propagating multiple sensitivities at once. Monitoring functionality gives users insight into memory use, while the expression template engine and highly optimized internal data structures give best-in-class performance.

This flexibility and rich feature set allow users to write extremely efficient AD implementations:

- Greeks are computed in seconds rather than hours, giving more responsive hedging at reduced hardware costs
- Desks have all first (and second) order Greeks rather than just a few, allowing better risk management
- Calibrations are significantly faster and more stable thanks to accurate gradient information

The bank in this Case Study chose NAG® dco/c++ and embedded it into their XVA platform. They also chose the NAG® AD Library for core analytics such as optimization, root finding, linear algebra and interpolation. The NAG® AD Library provides seamless integration with NAG® dco/c++, allowing backpropagation through NAG's huge collection of tried and tested numerical kernels. The NAG® AD Library also has a growing set of highly efficient symbolic adjoints covering root finding, solutions of linear systems, optimization and nearest correlation matrix problems (among others). These symbolic adjoints differentiate the mathematics, not the code, and give dramatic speedups over a naïve application of AD.

CASE STUDY

Elegant Distribution

Fast backpropagation alone is of little use if the underlying work distribution engine cannot handle the complexity. XVA naturally lends itself to task-based execution with each task performing a specific job (calibration, curve building, Monte Carlo, etc). Dependencies between tasks are expressed as a Directed Acyclic Graph (DAG). This maps well onto the mathematics, decomposing the problem into manageable pieces and making things easier, provided the work distribution engine can handle it. Yet many enterprise grid management systems (EGMS) struggle with this for three main reasons:

- DAGs are not first-class citizens, and the latency for task-scheduling can be significant
- It is incredibly difficult to schedule tasks between GPU or CPU and manage the load balance
- The EGMS are highly complex and often fragile, requiring constant attention and maintenance.

After evaluating several alternatives, the client selected Origami, NAG's solution for efficient and flexible task orchestration. Origami allows banks to avoid all the previously mentioned shortcomings: it is a lightweight, modern, and above all stable work distribution engine supporting a wide range of use cases.

Origami is easy to configure and deploy, and runs equally well on workstations, production systems, in-house grid, cloud, or any combination of these.

Quants can test for correct behaviour on their development machines before deploying to the grid, yielding faster development cycles and more stable releases. Tasks are created, scheduled, and executed with very low latency, crucial when full XVA revaluations can spawn hundreds of thousands of tasks, many of them quite small. Most importantly, Origami is a reliable platform requiring minimal maintenance once deployed.

Reducing Costs and Improving Risk Management

Automatic differentiation is a disruptive technology with the potential to cut computational costs by factors of 20x or more, while delivering huge amounts of accurate sensitivity data. The technology our client used to build their XVA engine can deliver substantial savings to your organisation as well. NAG works alongside clients to resolve tough technical problems (such as backpropagation through CUDA Monte Carlo codes), and our expertise helps our customers save money whilst creating smarter analytics. XVA desks are increasingly turning to NAG to help transform the way they operate, replacing the handful of sensitivities from expensive overnight bump-and-revaluation runs with rich first and second order sensitivity information. The result is a better understanding of the risks facing the bank, better tools to help manage that risk, and better service to customers.

Why not contact us to explore what NAG can do for you?

www.nag.com/reverse-mode-ad

About NAG

NAG provides industry-leading numerical software and technical services to banking and finance, energy, engineering, and market research, as well as academic and government institutions. World renowned for the NAG® Library - the most rigorous and robust collection of numerical algorithms available – NAG also offers Automatic Differentiation, Machine Learning, and Mathematical Optimization products, as well as world-class technical consultancy across HPC and Cloud HPC, code porting and optimisation, and other areas of numerical computing. Founded more than 50 years ago from a multi-university venture, NAG is headquartered in Oxford, UK with offices in the UK, US, EU and Asia.

NAG LTD
OXFORD, UK
www.nag.co.uk
+44 1865 511245

NAG INC
CHICAGO, USA
www.nag.com
+1 630 971 2337

NIHON NAG
OKYO, JAPAN
www.nag-j.co.jp
+81 3 5542 6311