

# Algorithmic differentiation for callable exotics: PV and XVA

Alexandre Antonov  
Numerix

Global Derivatives, Trading and Risk Management  
Barcelona

May 10, 2017

# Outline

- ▶ Introduction
- ▶ Algorithmic Differentiation for PV greeks: path-by-path Adjoint Differentiation
- ▶ Algorithmic Differentiation for exotic instruments with the Monte Carlo regression: Adjoint Differentiation vs Backward Differentiation (new method)
- ▶ Discussion on necessity of the regression differentiation
- ▶ Numerical experiments: XVA for a Bermudan swaption

Greeks calculation is complicated and time consuming operations in the financial software.

Here we concentrate on the *model* Greeks: price/XVA change w.r.t. the change of the model parameters

Our goal is the "*Rolls-Royce*" of the computational finance:

- ▶ compute greeks of multiple XVA's (CVA, DVA, FVA, COLVA etc) with a general collateral (CSA)
- ▶ for an instrument set containing all kinds of instruments:
  - ▶ "forward" (e.g. swaps, European options, Barrier options)
  - ▶ "backward" (e.g. Bermudan or American options)
- ▶ using MC methods including regressions

# Methods for greeks

## Methods for model greeks

- ▶ Bump-and-reprice (BnR)
- ▶ Payoff differentiation (the Adjoint Differentiation (AD) is a part of it)
- ▶ Likelihood (notably, Malliavin techniques, powerful but very bespoke)

In general, the *quality* of the BnR derivatives are equivalent to the payoff differentiation ones. The *speed* depends on multiple factors, mainly on the number of market bumps w.r.t. model parameters.

# Traditional Payoff Differentiation

A Traditional Payoff Differentiation goes *path-by-path*; it can be either direct (towards the result) or adjoint (back from the result):

- ▶ direct differentiation is efficient when the number of results (outputs) is more than the number of parameters (input)
- ▶ adjoint differentiation (AD) is efficient in the opposite case (more natural in finance)

The AD was brought into mathematical finance by Glasserman-Gilles in 2006 and widely spread after that (Capriotti, Andreasen and many others):

- ▶ The AD coding is considered as complicated
- ▶ It requires a lot of memory to write the tape
- ▶ However, its "theoretical" speed (number of multiplications) is impressive: *it does not depend of the number of parameters but is only 4-10 times slower than the function itself*

# AD logic

Using a simple example below we show why the AD speed is so high.

Introduce:

- ▶ Model parameters:  $N_\theta$ -dimensional vector

$$\theta = \{\theta_1, \dots, \theta_{N_\theta}\}$$

- ▶ Model states:  $N_X$  dimensional Markovian process in a model measure and expectation  $\mathbb{E}[\cdot]$

$$X(t) = \{X_1(t), \dots, X_{N_X}(t)\}$$

depending on the parameters  $X(t) = X(t, \theta)$  defined on a discrete set of timesteps by general Markov discretization scheme on a set of  $N_t$  simulation dates  $\{t_i\}_{i=1}^{N_t}$

$$X(t_{i+1}) = G(t_i, X(t_i), Z_i, \theta),$$

where  $Z_i$  is, for example, Gaussian random variable. The origin is at  $t_0$ .

Consider a general path-dependent payoff

$$Q(X) = Q(X(t_1), \dots, X(t_{N_t}))$$

and its value  $V(\theta) = \mathbb{E}[Q(X(\theta))]$  calculated using the Monte Carlo (MC) simulations

$$V(\theta) \simeq N_p^{-1} \sum_{p=1}^{N_p} Q(X(\theta)[p]),$$

where  $N_p$  is the number of MC paths and  $X(t)[p]$  is the state realization of the  $p$ th path.

The goal  $\rightarrow$  calculate the price derivative over all params  $\{\theta_n\}$ :

$$\frac{\partial V(\theta)}{\partial \theta} \simeq \frac{1}{N_p} \sum_{p=1}^{N_p} \frac{\partial Q(X)[p]}{\partial \theta}$$

*Remark.* We calculate the underlying derivatives  $\partial Q[p]/\partial \theta$  path-by-path and for notational brevity, omit the path index.

The AD calculation is based on the **backward substitution** of the simulation step into the payoff:

$$\begin{aligned} Q(X(t_1), \dots, X(t_{N_t})) &= \\ &= Q(X(t_1), \dots, X(t_{N_t-1}), X(t_{N_t}, G(t_{N_t-1}, X(t_{N_t-1}), Z_{N_t-1}, \theta)) \\ &= \dots \\ &= Q_i(X(t_1), \dots, X(t_i); Z_i, \dots, Z_{N_t-1}; \theta), \end{aligned}$$

where we have denoted by  $Q_i$  the payoff value for which we have backwardly substituted the states until its latest direct dependence on  $X(t_i)$ .

**Our goal** → find the payoff dependence on the states for which we have substituted all of the  $\theta$  dependencies:

$$Q_0(Z_1, \dots, Z_{N_t-1}; \theta)$$

and its derivatives over the parameters  $\partial Q_0 / \partial \theta$

The recursive relationship between  $Q_i$  and  $Q_{i+1}$ ,

$$\begin{aligned} Q_i(X(t_1), \dots, X(t_i); Z_i, \dots, Z_{N_t-1}; \theta) \\ = Q_i(X(t_1), \dots, X(t_{i-1}), G(t_{i-1}, X(t_{i-1}), Z_{i-1}, \theta); Z_i, \dots, Z_{N_t-1}; \theta) \\ = Q_{i-1}(X(t_1), \dots, X(t_{i-1}); Z_{i-1}, \dots, Z_{N_t-1}; \theta), \end{aligned}$$

is the key to expressing the payoff derivatives *backwardly*:

$$\begin{aligned} \frac{\partial Q_{i-1}}{\partial \theta} &= \frac{\partial Q_i}{\partial \theta} + \frac{\partial Q_i}{\partial X(t_i)} \frac{\partial G(t_{i-1}, X(t_{i-1}), Z_{i-1}, \theta)}{\partial \theta} \\ \frac{\partial Q_{i-1}}{\partial X(t_{i-1})} &= \frac{\partial Q_i}{\partial X(t_{i-1})} + \frac{\partial Q_i}{\partial X(t_i)} \frac{\partial G(t_{i-1}, X(t_{i-1}), Z_{i-1}, \theta)}{\partial X(t_{i-1})} \end{aligned}$$

We call these two equations **the AD recursion**.

The AD is a reverse scheme w.r.t. the calculation process → it requires storage of the following information during the model simulation and instrument pricing

- ▶ *Simulation info* (simulation tape): derivatives of the discretization function:

$$\frac{\partial G(t_i, X(t_i), Z_i, \theta)}{\partial \theta} \quad \text{and} \quad \frac{\partial G(t_i, X(t_i), Z_i, \theta)}{\partial X(t_i)},$$

- ▶ *Instrument info* (instrument tape)  
partial derivatives of the payoff,

$$\frac{\partial Q_i}{\partial X(t_{i-1})},$$

which is a vector of length  $N_X$  per path per time step.

For any fixed payoff  $Q$ , AD requires the reverse propagation.

Having calculated and stored the *tapes* we go backward in time, calculating the derivatives.

Cost of the AD operations:

- ▶ The derivative calculations  $\partial Q_i / \partial X(t_i)$  cost  $N_x^2$  multiplications per path and per timestep.
- ▶ The derivative calculations  $\partial Q_i / \partial \theta$  cost  $O(1)N_x$  multiplications per path and per timestep for **all** parameters.

*Remark.* Our example is a "manual" AD; there exists softwares permitting an automatic AD (with different degree of efficiency) where the simulation and pricing are not separated but the prices treated algorithmically as general function of the parameters

```
Price=FUNCTION(theta1, ..., thetaN)
```

# The AD for the callable exotics

- ▶ The speed of greeks calculation becomes important (XVA hedging for exotics portfolios, initial margin, FRTB etc.)
- ▶ The regression (least-square MC, AMC) is the key numerical method for exotic deals; it is one of the most complex and time-consuming numerical methods.
- ▶ It introduces path interdependencies that compromise the path-by-path valuation: the AD becomes difficult. That is why people often avoid the regression if they can.
- ▶ In A<sup>2</sup> (2016) and AIKMM (2016) we have found a universal method that can be applied uniformly across all exotic deals for both price and XVA greeks with full regression mechanism.
- ▶ We will illustrate the underlying mathematical logic on example of a Bermudan swaption.

- ▶ Callable instruments pricing procedure is encoded in so called pricing script.
- ▶ Its main component is a *continuation value* (CV). We denote it as  $V(t)$  at time  $t$ .
- ▶ The CV is a certain function of the model state variables at time  $t$ . Financially,  $V(t)$  is a *hold* value: it is the price of holding an option at time  $t$  with all possible payments and exercises taking place thereafter.
- ▶ A simple CV example is a leg with a single unit payment at time  $T$ . More complicated examples include a swap value or a swaption holding value. For a European option the CV is a discounted conditional expectation of the payoff.

## Bermudan swaption

- ▶ Consider a Bermudan swaption giving the right to enter into a swap on exercise dates  $T_i$ .
- ▶ The swap pays a *generalized cashflow*  $c_j$  at date  $T_j$  (for simplicity coinciding with the exercise dates). The cashflows can be, for example, fix or floating rates. The cashflows are known analytically as functions of states and parameters.
- ▶ Let  $V(t)$  and  $S(t)$  denote the CVs of the swaption and the swap
- ▶ The swaption's backward pricing is performed using two repeated steps encoded in a pricing script



- ▶ The first step: **transformations** at  $T_j$

$$V(T_j^-) = \max(V(T_j), S(T_j))$$

$$S(T_j^-) = S(T_j) + c_j$$

where  $T^- = T - \varepsilon$

- ▶ The second step: **propagation** (the discounted conditional expectation) between the instrument dates:

$$S(t) = N(t) \mathbb{E} \left[ \frac{S(T)}{N(T)} \middle| \mathcal{F}_t \right] \quad \text{and} \quad V(t) = N(t) \mathbb{E} \left[ \frac{V(T)}{N(T)} \middle| \mathcal{F}_t \right]$$

where  $N(t)$  is the model numeraire.

Often, the CV's  $S(t)$  and  $V(t)$  are certain function of the *model states* (driving factors)  $X(t)$ .

A general instrument contains a set of CVs  $V_j$  that are transformed by means of  $+$ ,  $-$ ,  $*$ ,  $/$ ,  $\max$ , step etc.

# Regression

In practice, for conditional expectations, we use the MC regression (a.k.a. least-squares MC, AMC). Each CV is presented as a vector of path values  $V(t) = \{V(t)[p]\}_{p=1}^{N_p}$ .

Fix two times  $t_1$  and  $t_2$  and denote  $V(t_i) = V^{(i)}$  and  $X(t_1) = X$ . For the numerical calculation of the conditional expectation

$$V^{(1)}(x) = \mathbb{E} \left[ V^{(2)} \mid X = x \right],$$

we need to fix basis functions  $\{\phi_n(X)\}_{n=1}^{N_b}$  and determine coefficients  $\beta_n$  which minimize the expectation

$$\chi^2 = \mathbb{E} \left[ \left( V^{(2)} - \sum_n \beta_n \phi_n(X) \right)^2 \right],$$

*Remark.* For simplicity we ignore numeraire in the expectations (See AIKMM (2016) for complete formulas.)

The unconditional expectation  $\chi^2$  is calculated over the MC paths:

$$\chi^2 \simeq \frac{1}{N_p} \sum_{p=1}^{N_p} \left( V^{(2)}[p] - \sum_n \beta_n \phi_n(X[p]) \right)^2.$$

Abusing notations we denote by  $V^{(1)}[p]$  the  $p$ th path of results for the MC regression on our basis  $\phi_n(X)$ .

Such stochastic variable is an approximation to the exact conditional expectation. The errors come from two sources:

- ▶ Incomplete basis (the main error)
- ▶ MC statistical error

In all of the regression-based methods, the choice of the basis is the key.

Below, we treat all of our equations in the MC sense assuming that

- 1) averages are MC sums  $\mathbb{E}[X] = (1/N_p) \sum_p X[p]$
- 2) conditional averages are results of the corresponding regressions.

Minimizing  $\chi^2$  yields the set of equations which can be easily solved:

$$V^{(1)}[p] = \frac{1}{N_p} \sum_{nm,p'} \phi_n(x[p]) C_{nm}^{-1} \phi_m(x[p']) V^{(2)}[p'].$$

where

$$C_{nm} = \mathbb{E} [\phi_n(X) \phi_m(X)]$$

is the basis function covariance matrix.

The linear operation  $V^{(2)} \rightarrow V^{(1)}$  “mixes the paths”, i.e. a value  $V^{(1)}[p]$  will depend on all the paths of  $V^{(2)}[p']$  and  $X[p']$ .

This leads to derivatives of the regression containing *cross-terms* (result at path  $p$  depend on states for all paths)



$$\frac{\partial V^{(1)}[p]}{\partial V^{(2)}[p']} = \frac{1}{N_p} \sum_{nm} \phi_n(x[p]) C_{nm}^{-1} \phi_m(x[p'])$$

$$\frac{\partial V^{(1)}[p]}{\partial X[p']} = \delta_{pp'} V^{(1)'}[p] + \frac{1}{N_p} \sum_{nm} \phi_n(X[p]) C_{nm}^{-1}$$

$$\times \left\{ \phi'_m(X[p']) (V^{(2)}[p'] - V^{(1)}[p']) - \phi_m(X[p']) V^{(1)'}[p'] \right\}$$

where  $V^{(1)'}[p]$  is the derivative of the regression result

$$V^{(1)'}[p] = \sum_n \beta_n \phi'_n(X[p])$$

and  $\phi'_n(x)$  is the derivative of  $\phi_n(x)$  over its argument.

*Remark.* The derivative matrices size is large ( $N_p \times N_p$ ) and they heavily mix the paths. However, their rank is equal to the number of basis functions  $N_b$ , which is often much smaller than  $N_p$ .

## Two ways to proceed for PV/XVA greeks

In AIKMM (2016) we have studied the algorithmic differentiation for exotics pricing where we have selected two ways:

- ▶ Standard AD
  - the method works directly with the PV derivatives and requires the instrument tape, a place in memory where we write all the pricing/XVA operations. This memory consumption is known to be very high. Using the adjoint pass we use the tape "playback" to perform the differentiation.
- ▶ New Backward Differentiation
  - The method delivers continuation values (CV) sensitivities.
  - The method does not require the instrument tape: the differentiation is done at the same time as the pricing.

The standard AD for exotics was also treated in the independent research by Capriotti et al. (2016).

## New Backward Differentiation (BD) for PV greeks

- ▶ This approach to the algorithmic differentiation works with *continuation values* instead of the final payoffs as in the usual AD.
- ▶ This allows us to avoid the instrument tape: the differentiation is done *during* the backwards pricing. We note that the simulation tape is still present.
- ▶ Absence of the instrument tape saves a lot of memory as well as the time needed to access it.
- ▶ Moreover, it makes the code simpler and easier to debug.
- ▶ Finally, we can obtain greeks for all of the CVs in a pricing script, e.g., swaptions, underlying swaps, legs, etc., in a single pass of the BD (the usual AD procedure must be rerun for every new payoff)

## Backwards propagation details

The backwards propagation  $V(t) = \mathbb{E}[V(T)|\mathcal{F}_t]$  is the key part of the algorithm. It transforms the info at time  $T$

$$\{V(T)[p], \partial V(T)[p]/\partial X(T)[p'], \partial V(T)[p]/\partial \theta\}$$

into that at time  $t$ ,

$$\{V(t)[p], \partial V(t)[p]/\partial X(t)[p'], \partial V(t)[p]/\partial \theta\}$$

*Example.*

$$\underbrace{\frac{\partial V(t)[p]}{\partial \theta}}_{t \text{ info}} = \sum_{p''} \underbrace{\frac{\partial V(t)[p]}{\partial V(T)[p'']}}_{\text{from the regression}} \underbrace{\frac{\partial V(T)[p'']}{\partial \theta}}_{T \text{ info}}$$

All the technical details are in A<sup>2</sup> (2016).

## Generalization to the XVA

So far we have treated the *price greeks*. Now we present the more complicated subject of the *XVA greeks*.

This will require so called future values (FV) which we denote with small letters  $v(t)$ . Their main difference with the CV's is that the FV takes into account possible exercises before the time  $t$  whereas the CV does not.

For a general collateral, a general **XVA** is defined on a set of observation dates  $\{s_i\}$

$$XVA = \sum_i \mathbb{E} [F_i(u(s_1), \dots, u(s_i)) / N(s_i)] w_i$$

where deterministic function  $F_i(u_1, \dots, u_i)$  is a generalized portfolio exposure and  $w_i$  are some weights.

*Example.* One-sided CSA **CVA** =  $\sum_{s_i} \mathbb{E} [v(s_i)^+ / N(s_i)] w_i$

## Example: Bermudan swaption

The swap pays cashflows  $c_j$  at a date  $T_j$  for  $j = 1, \dots, M$   
→ the swaption PV is built with payments subjected to the exercise conditions

$$V = \mathbb{E} \left[ \sum_{j=1}^M \frac{\mathcal{I}(T_j) c_j}{N(T_j)} \right]$$

where the global exercise indicator  $\mathcal{I}(T_j)$  equals one if we have entered into the swap *before* the payment date  $T_j$  and zero otherwise.

The global exercise indicator can be computed given the local indicators

$$\mathcal{C}_i = 1_{S(T_i) > V(T_i)}$$

defined on exercise dates  $T_i$  in terms of the swaption/swap continuation values.

The global exercise indicator for  $T_j \leq t < T_{j+1}$  reads

$$\mathcal{I}(t) = 1 - \prod_{i=1}^j (1 - \mathcal{C}_i).$$

Now we can define the swaption *future value* (FV) for a given date  $t$ : it is a conditional expectation of all payments on and after  $t$

$$v(t) = N(t) \mathbb{E} \left[ \sum_{\substack{j=1, \\ T_j \geq t}}^M \frac{\mathcal{I}(T_j) c_j}{N(T_j)} \middle| \mathcal{F}_t \right],$$

One can prove that the swaption FV can be expressed through the CV's and global exercise indicators

$$v(t) = V(t) (1 - \mathcal{I}(t)) + S(t) \mathcal{I}(t)$$

In spite of its complicated and "bespoke" structure there is an algorithmic way to calculate it (AIM (2015)).

## BD for the FV calculations

- ▶ Combining the BD with the Algorithmic Calculation for the future values we come up with the BD for the XVA
- ▶ For each CV and associated FV's we calculate their derivatives over the parameters following the differentiation rules
- ▶ The result is a set of derivatives

$$\frac{\partial v(s_i)[p]}{\partial \theta_k}$$

for all the paths  $p$ , parameters  $k$  and observations  $i$ .

- ▶ Then, the calculated derivatives go to the XVA formula.
- ▶ We stress that the BD for the XVA does not contain the instrument tape

Details can be found in AIKMM (2016).

## BD vs. AD for XVA calculations

- ▶ SPEED

Both methods have comparable time: much faster than BnR.

- ▶ TAPE AND STORAGE

The BD does not use the instrument tape. Both BD and AD have comparable storage requirements.

- ▶ MULTIPLE XVA's

The BD delivers model greeks for all of the XVA's (CVA, DVA, FVA etc). The AD should be rerun for each new XVA.

- ▶ CODING/DEBUGGING

The AD coding is complicated: it sometimes requires the library rewriting. The tape adds a lot of complexity to the coding, debugging, and maintenance. The BD implementation is much simpler and compatible with the price scripting.

- ▶ COMBINATION WITH BUMP-AND-REPRICE

The BD can be combined with the BnR (but not for the AD).

## Do we always need derivatives of the regression?

Quite often we know analytical dependence of the cashflows  $c_k$  from the parameters

$$\frac{\partial c_k[p]}{\partial \theta}$$

We suppose that the derivatives are explicit (we have substituted all the states dependence).

**Question:** can we use the derivatives of the cashflows to calculate XVA's or need to differentiate the results of the regression?

## PV case

Consider first our Bermudan where the exercise indicators form a surface in the state space for each exercise date. This surface is *optimal* for the Bermudan swaption price

$V = \mathbb{E} \left[ \sum_{j=1}^M \mathcal{I}_{j-1} \frac{c_j}{N(T_j)} \right]$  i.e. an infinitesimal "bump" of the indicators as functions of parameters leads to zero

$$\mathbb{E} \left[ \sum_{j=1}^M \frac{\partial \mathcal{I}_{j-1}}{\partial \theta} \frac{c_j}{N(T_j)} \right] = 0.$$

This immediately gives us the PV Greek

$$\frac{\partial}{\partial \theta} V = \mathbb{E} \left[ \sum_{j=1}^M \mathcal{I}(T_j) \frac{\partial}{\partial \theta} \left( \frac{c_j}{N(T_j)} \right) \right]$$

which requires only differentiation of the cashflows and the numeraire.

Thus, the optimal exercise instruments will not depend on the exercise derivatives and the differentiation of the conditional exercises can be avoided.

This rule can be extended to other (not path-dependent) backward<sup>1</sup> instruments.

To prove it we associate with any CV  $V(t)$  its “noisy derivative”  $D_\theta^V(t)$  such that its conditional expectation equals to the exact derivative of the CV

$$\mathbb{E} \left[ D_\theta^V(t) \mid \mathcal{F}_t \right] = \frac{\partial V(t)}{\partial \theta}$$

*Remark.* The noisy derivative would be exact if we were differentiating the regression (but we want to avoid it).

---

<sup>1</sup>Or written in the backward way, like barriers etc.

# Algorithm

- When the CV is modified, the noisy derivatives will be modified according to the differentiation rules, e.g.,

$$V(T_k^-) = \max(V(T_k), S(T_k))$$

↓

$$D_\theta^V(T_j^-) = (1 - \mathcal{C}_j)D_\theta^V(T_j) + \mathcal{C}_j D_\theta^S(T_j)$$

or

$$S(T_k^-) = S(T_k) + c_k \Rightarrow D_\theta^S(T_k^-) = D_\theta^S(T_k) + \frac{\partial c_k}{\partial \theta},$$

where  $\partial c_k / \partial \theta$  is a real derivative of the analytically available cashflows.

- ▶ Conditional expectations for the CVs induce a simple discounting for noisy derivatives, i.e.,

$$D_\theta^V(t^-) = D_\theta^V(T^-) \frac{N(t)}{N(T)} + V(T^-) \frac{\partial}{\partial \theta} \left( \frac{N(t)}{N(T)} \right).$$

- ▶ At the origin, an expectation of the noisy derivative gives the correct CV derivative

$$\frac{\partial V(0)}{\partial \theta} = \mathbb{E} \left[ D_\theta^V(0) \right]$$

Limitations of the approach above are related with **different time** combinations of the CV's, i.e. path-dependence of the instrument (e.g. the following CV update  $V(T_2^-) = V(T_2) 1_{U(T_1) \geq 0}$ )

**Conclusion on PV greeks.** For any non path-dependent backward instrument, one can avoid the differentiation of the conditional expectations.

- ▶ Consider a simple credit valuation adjustment (CVA) with a deterministic hazard rate  $\lambda(t)$  and a deterministic collateral  $C(t)$ .
- ▶ Suppose that the only element of portfolio is our Bermudan.

The **total** instrument price

$$\hat{V} = \mathbb{E} \left[ \sum_{j=1}^M \mathcal{I}_{j-1} \frac{c_j}{N(T_j)} \right] - \mathbb{E} \left[ \int_0^T dt \tilde{\lambda}(t) \frac{(v(t) - C(t))^+}{N(t)} \right]$$

where  $v(t)$  is a future value (FV) of the Bermudan and  
 $\tilde{\lambda}(t) = \lambda(t) e^{-\int_0^t ds \lambda(s)}$ .

The last term is called CVA.

The FV  $v(t)$  is path-dependent  $\Rightarrow$  one *cannot* find its noisy derivative having information about the cashflow derivatives. However, as shown in AIKMM (2016), we can transform each fixed time  $t$  term

$$\mathbb{E} \left[ \frac{(v(t) - C(t))^+}{N(t)} \right]$$

into a backward instrument (effectively making the instrument non-path-dependent) and *get rid of the differentiation of conditional expectations*.

The computational cost is high: for the CVA we get

# of CVA discretization time-steps  $\times$  # of exercises

of extra regressions.

Such procedure may be much slower than the regression differentiation!

## Collateral influence

Consider a portfolio of vanillas with the noisy derivatives coinciding with the cashflow derivatives, i.e.  $\mathbb{E}[D_\theta^\nu(t) | \mathcal{F}_t] = \partial v(t) / \partial \theta$ .

Establish conditions on the *collateral* when the cashflow derivatives are sufficient for the XVA Greeks.

**The simplest case:** collateral as a deterministic function of the portfolio value at the same time, i.e.  $C(t) = C(v(t))$ .

Derivatives of the loss function  $f(v) = (v - C(v))^+$   $\Rightarrow$

$$\begin{aligned}\frac{\partial \mathbb{E}[f(v(t))]}{\partial \theta} &= \mathbb{E}\left[\frac{\partial f(v(t))}{\partial \theta}\right] = \mathbb{E}\left[f'(v(t)) \frac{\partial v(t)}{\partial \theta}\right] \\ &= \mathbb{E}\left[f'(v(t)) D_\theta^\nu(t)\right]\end{aligned}$$

*Conclusion.* A Greek for a single time average of the vanilla portfolio FV depends only on cashflow derivatives (see also Andreasen (2014) for  $f(x) = x^+$ ).

- ▶ For a multi-time dependence of a general XVA

$$\mathbb{E}[f(v(t_1), \dots, v(t_n))]$$

corresponding to a general collateral with non-zero minimum transfer amount and threshold such property is not valid.

- ▶ However, one can employ a *branching* diffusion as was first suggested in Andreasen (2014) for  $f = (v(t_{n-1}) - v(t_n))^+$ .
- ▶ If we have a general path-dependence for the XVA, the branching logic becomes cumbersome ( $n$  re-simulations for timestep  $t_n$ ), slow and hard to control on the script level.

*Thus, for a general XVA with a general collateral one should differentiate the conditional expectation.*

## Summing up

Conditions when the cashflow derivatives are not sufficient (or it is too slow to avoid differentiation of conditional expectation) to calculate XVA Greeks:

- ▶ The portfolio contains Bermudan/American options which are exercised in the usual risk-free way
- ▶ The collateral is generally dependent on multiple time horizons

For other cases, mainly representing vanilla portfolios with a simple collateral, it is sufficient to use cashflow derivatives for XVA Greeks.

## Numerical experiments

- ▶ We numerically demonstrate the efficiency of the backward algorithmic differentiation using the HW1F model with constant parameters: the volatility is 2%, the reversion is 5%, and the rate is 2%.
- ▶ The exotic instrument in hand is a 10-year Bermudan ATM swaption with coinciding quarterly fixing and exercise dates from 0.25 years to 10.25 years. The payment dates are shifted forward by 0.25 years.
- ▶ The swaption price is 0.107 for a unit notional.
- ▶ We use 12 observation dates per year: **126** in total. This results in **252 regressions** both for swap and swaption.

- ▶ We cover two experiments: the first one is related to the PV Greeks while the second one with the XVA Greeks. We calculate sensitivities to different buckets of the volatility (Vega) and forward rates of the yield curve (Delta).
- ▶ We bump these parameters either on the annually spaced intervals (11 buckets) or on the quarterly intervals (43 buckets).
- ▶ For both experiments we perform Monte-Carlo simulation **5,000** high-quality, low-discrepancy paths with 126 time steps. The conditional expectation is calculated with regression on a set of **17** smooth basis functions similar to Carriere (1996).
- ▶ We present timing of our new method, alternative to the AD, called Backward Differentiation (**BD**) vs. bump-and-reprice (**BnR**) and acceleration results (greeks profiles as well as convergence analysis can be found in AIKMM (2016))

# of Vols	# of Rates	Pricing	BD	BnR	BnR/BD	BD/Pricing
11	11	0.7	1.3	15.3	12.2	1.8
11	43	0.7	1.5	36.9	24.4	2.2
43	43	0.7	2.5	59.2	23.5	3.7

Table: Time (sec) for **Pricing Greeks** with regression differentiation

# of Vols	# of Rates	Pricing	BD	BnR	BnR/BD	BD/Pricing
11	11	0.7	0.8	15.2	20.2	1.1
11	43	0.7	0.8	38.3	46.5	1.2
43	43	0.7	1.0	58.6	57.2	1.5

Table: Time (sec) for **Pricing Greeks** without regression differentiation

# of Vols	# of Rates	XVA	BD	BnR	BnR/BD	BD/XVA
11	11	1.8	3.9	39.4	10.1	2.2
11	43	1.8	6.8	97.1	14.3	3.8
43	43	1.8	11	155.1	14.1	6.1

Table: Time (sec) for **XVA Greeks** with regression differentiation

# Observation

According to the AD rule of thumb (the AD is 4-10 times slower than the pricing) the BD performs favorably:

- ▶ Pricing greeks
  - up to over 20 times faster than the BnR with regression differentiation, and almost up to 60 times faster without it.
- ▶ XVA greeks
  - up to over 15 times faster than the BnR

- ❑ Andreasen, J. CVA on an iPad Mini. Global Derivatives, ICBI, Amsterdam, 2014.
- ❑ Antonov, A., Issakov, S., and Mechkov, S. Backward induction for future values. *Risk Magazine* (January 2015), 92–97.
- ❑ Antonov, A., Algorithmic Differentiation for Callable Exotics (September 2016), SSRN preprint
- ❑ Antonov, A., Issakov, S., Konikov, M., McClelland, A., and Mechkov, S., PV/XVA Greeks for Callable Exotics by Algorithmic Differentiation (December 2016), SSRN preprint
- ❑ Capriotti, L., and Giles, M. Adjoint greeks made easy. *Risk Magazine* (September 2012), 96–102.

- Capriotti, L., Jiang, Y. and Macrina, A. AAD and least squares Monte Carlo: Fast Bermudan-style options and XVA Greeks (September 2016), SSRN preprint
- Carriere, J., Valuation of the early-exercise price for options using simulations and nonparametric regression (1996) *Insurance: Mathematics and Economics*, 19:19(30)
- Giles, M., and Glasserman, P. Smoking adjoints: fast Monte Carlo Greeks. *Risk Magazine* (January 2006), 88–92.