Convexité, robustesse et apprentissage profond pour l'évaluation et la couverture d'options

Vincent Lemaire & Gilles Pagès

LPSM-Sorbonne-Université

(Labo. Proba., Stat. et Modélisation)

Webinaire Summit Maths-fi & Assurance

17 février 2021

$$\heartsuit \quad \overset{?}{\leftrightsquigarrow} \quad \heartsuit$$

1D dynamics of a traded asset

- $(X_t^x)_{t\in[0,T]}$ martingale diffusion representative of the risk-neutral quotation of a traded asset.
  - Typically, Black-Scholes model (1973),

$$X_t^x = x + \int_0^t X_s^x \sigma dW_s \longleftrightarrow X_t^x = x\, e^{\sigma W_t - \frac{\sigma^2}{2} t},\ x > 0,\ t \in [0, T].$$

  - Local volatility model $[\sigma(x) = \vartheta x^{-1/2}$ for CEV model$]$

$$X_t^x = x + \int_0^t X_s \sigma(X_s^x) dW_s.$$

  - Stochastic volatility model [Heston $\vartheta(y) = \sqrt{y}$ (1993)]

$$X_t^x = x + \int_0^t X_s^x \sigma(V_s^v) dW_s, \quad V_t^v = v + \int_0^t (a - \kappa V_s^v) ds + \vartheta(V_s^v) dB_s,\ V_0 = v$$

  with $\langle W, B\rangle_t = \rho t,\ \rho \in [-1, 1]$.
  - Loc-Stoch. volatility model [like SABR model]

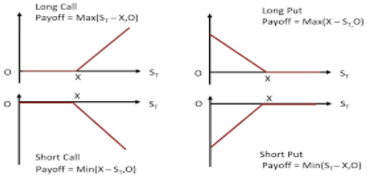
$$X_t^x = x + \int_0^t X_s^x \sigma(X_s^x, V_s^v) dW_s \quad + \quad \text{stoch. vol.}$$

Convex vanilla payoffs

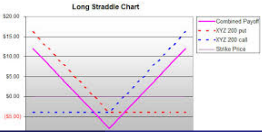
- European option with maturity $T > 0$ and payoff $\varphi(X_T^x) \geq 0$ has price and δ-hedge given by

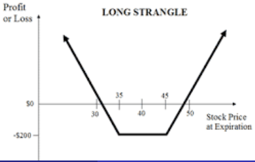
$$\mathbb{E}\,\varphi(X_T^x) \quad \text{and} \quad \partial_x \mathbb{E}\,\varphi(X_T^x).$$

- Call & Put payoffs (gross profits)



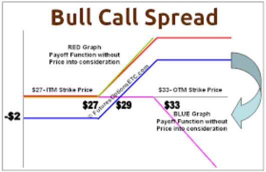
- Straddle & strangle payoffs (gross and net profits)

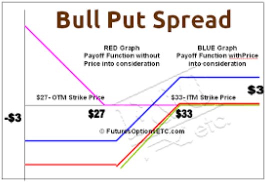




Non convex payoffs ?

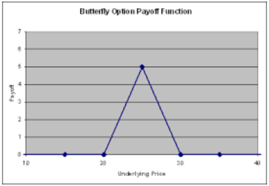
- Call & Put spread payoffs





$$(X_T - K_1)^+ - (X_T - K_2)^+ \qquad \& \qquad (K_1 - X_T)^+ - (X_T - K_2)^+$$

- Butterfly (difference of two spread options)



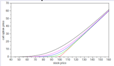
$$\varphi(X_T) = (X_T - K_1)^+ - 2\Big(X_T - \frac{K_1 + K_2}{2}\Big)^+ + (X_T - K_2)^+$$

- In fact, most payoffs are linear combination of convex payoffs. . .

What does theory say ?

Theorem (Bergman et al, 1996)

([a]) For local volatility models for every convex payoff function $\varphi : \mathbb{R} \to \mathbb{R}_+$

$x \mapsto \mathbb{E}\,\varphi(X_T^x)$ is also convex. (e.g. Black-Scholes Call )

[a] General Properties of Option Prices Y. Z. Bergman, B. D. Grundy and Z. Wiener, *J. of Finance*, **51**(5):1573-1610, 1996.

- Intuitive, because of Black-Scholes' world, but not always true !
- Stoch. volatility models: obviously true since by one integration

$$\mathbb{E}\,\varphi(X_T^x) = \mathbb{E}\,\hat{E}\varphi\Big(x.exp\Big(\int_0^T \sigma(V_S)dW_s - \frac{1}{2}\int_0^T \sigma^2(V_s)ds\Big)\Big)$$

- But if Loc-Vol and/or V_t depends on X_t or multi-asset payoffs (basket, BoC, etc) may fail. . .

Theorem (Lemaire-P., 2021)

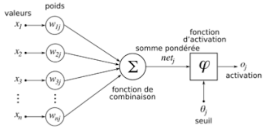
If a general $\mathbb{R}^d$-valued martingale diffusion $(Y_t^x)_{t\in[0,T]}$satisfies

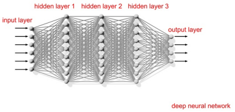
$$dY_t^x = x + \int_0^t \vartheta(Y_s^x)dW_s,$$

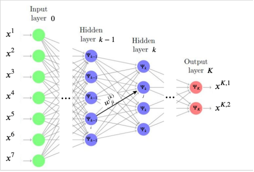
with W q-dim Brownian motion, and $x \mapsto \vartheta(x)$ is convex for the (matrix partial) order $A \preccurlyeq B$ if $BB^* - AA^* \geq 0$ then, for every $\mathbb{R}_+$-valued convex functional Φ,

$$x \mapsto \mathbb{E}\,\Phi(Y^x) \quad \text{is also convex.}$$

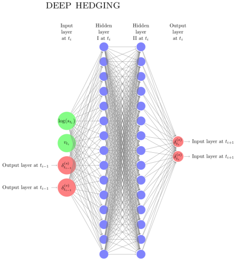
- Challenge ImageNet in 2012: G. Hinton-Y. Le Cun-Y. Bangio make a breakthrough with a Deep Neural network

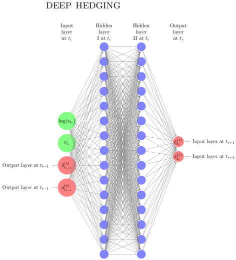


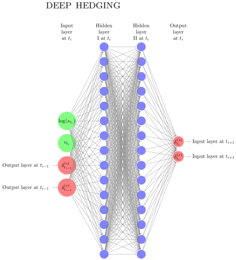




- JPMorgan (Bühler, et al.) et ETH Zürich (J. Teichmann) in 2017 propose Deep Hedging ([1]) to compute (price and) hedge of an option with a neural network.



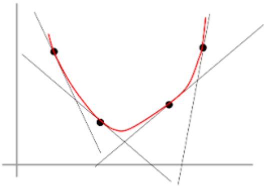


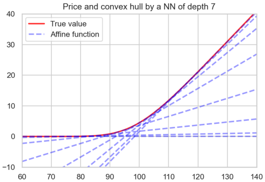


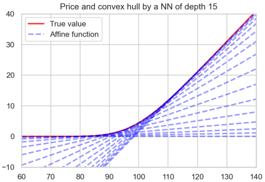
[1] Bühler, H.; Gonon, L.; Teichmann, J.; Wood, B. Deep hedging. *Quant. Finance*, **19**:1271-1291, 2019.

An alternative approach: ICNN

- (How to) Find an architecture that produces systematically a convex function of the input
$$x \mapsto ICNN(x, \mathbf{w}) \text{ is convex}$$
whatever the hyper-matrix of weights $\mathbf{w}$ is?
- Input Convex Neural Network







- Starting Idea [2] If $f : \mathbb{R}^d \to \mathbb{R}$ convex

$$f(x)\left\{ \begin{array}{l} \simeq \\ \geq \end{array} \max_{k=1:n} \left[f(x_k) + \langle \nabla f(x_k) | x - x_k \rangle \right] \right.$$

[2] B. Amos, L. Xu, J. Zico Kolter, Input Convex Neural Networks, ICML'17, 2017.

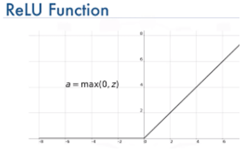
ICNN II

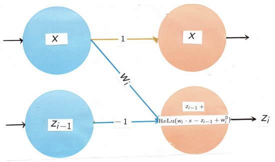
Design an appropriate neural network that emulates the convex functions

$$x \longmapsto \max_{k=1:n}\big[\langle w_k, x\rangle + w_k^0\big], \qquad (w_k^0, w_k) \in \mathbb{R}^{d+1}.$$

- Method 1([3]): change the connectivity of the network $\Longrightarrow$ specific learning script.
- Method 2: Set $z_0 = x$ and set z_{i-1} output of layer $i-1$

$$\max\big(z_{i-1}, \langle w_i \cdot x\rangle + w_i^0\big) = z_{i-1} + \mathrm{Relu}\big(\langle w_i \cdot x\rangle - z_{i-1} + w_i^0\big).$$





- Method 2 smooth: Replace $\mathrm{Relu}(x) = \max(x, 0)$ (Rectified Linear Unit) by a convex smooth (C^∞) approximation.

[3] B. Amos, L. Xu, J. Zico Kolter, Input convex Neural networks, ICML'17, 2017.

Loss functions

- Suggest to solve the problem(s)
  - On a compact sets e.g.

$$\mathcal{E}_K(f,n) := \min_{(w_k^0,w_k),k=1:n}\left[\sup_{|x|\le K}\left|f(x) - \max_{k=1:n}\varphi_k(x)\right|\right].$$

where $\varphi_k(x) := \langle w_k, x\rangle + w_k^0$ are affine functions ($(w_k^0, w_k) \in \mathbb{R}^{d+1}$) or

  - On the whole $\mathbb{R}^d$ with respect to a finite measure μ (empirical measure on a (simulated ?) dataset)

$$\mathcal{E}_{r,\mu}(f,n) = \min_{(w_k^0,w_k),k=1:n}\int_{\mathbb{R}^d}\left[\left|f(x) - \max_{k=1:n}\varphi_k(x)\right|^r\right]\mu(dx).$$

with $r = 1$ or $r = 2$.

Loss functions with penalization

- Suggest to solve the problem(s)
  - On a compact sets e.g.

$$\mathcal{E}_K(f,n,\varepsilon) := \min_{(w_k^0,w_k),k=1:n}\left[\sup_{|x|\le K}\Big|f(x)-\max_{k=1:n}\varphi_k(x)\Big| + \frac{1}{\varepsilon}\sum_{k=1}^{n}\Big(f(x)-\varphi_k(x)\Big)^{-}\right].$$

where $\varphi_k(x) := \langle w_k, x\rangle + w_k^0$ are affine functions ($(w_k^0, w_k)\in\mathbb{R}^{d+1}$) or

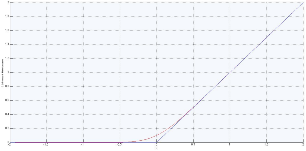
  - On the whole $\mathbb{R}^d$ with respect to a finite measure μ (empirical measure on a (simulated ?) dataset)

$$\mathcal{E}_{r,\mu}(f,n,\varepsilon) = \min_{(w_k^0,w_k),k=1:n}\int_{\mathbb{R}^d}\left[\Big|f(x)-\max_{k=1:n}\varphi_k(x)\Big|^r + \frac{1}{\varepsilon}\sum_{k=1}^{n}\Big(f(x)-\varphi_k(x)\Big)^{-}\right]\mu(dx).$$

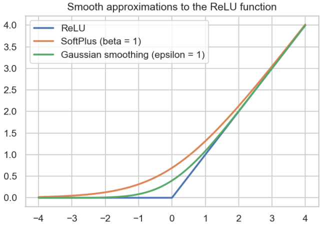
Probabilistic ε-smooth Relu

- Let $\zeta \sim \mathcal{N}(0,1)$.

$$\mathrm{Relu}_\varepsilon(x) := \mathbb{E}\,\mathrm{Relu}(x+\varepsilon\zeta) = x\mathbb{P}(\zeta < \tfrac{x}{\varepsilon}) + \varepsilon\frac{e^{-\frac{x^2}{2\varepsilon^2}}}{\sqrt{2\pi}}.$$



- Close to soft max (but most likely more adapted to option pricing/hedging)



Theoretical bounds

- Back to $w_k = \nabla f(x_k)$ and $w_k^0 = f(x_k) - \langle \nabla f(x_k), x_k \rangle$. Apropriate of $x_1, \ldots, x_n$ "guided" by covering numbers and $L^1(\mu)$-optimal quantization theory yields

Theorem (Balasz et al. 2015, Lemaire-P.2021)

(a) ([a]) If $f : B_{\mathbb{R}^d}(0, K) \to \mathbb{R}$ is convex and ∇f is bounded then

$$\mathcal{E}_K(f, n) = O(n^{-2/d}).$$

(b) If $f : \mathbb{R}^d \to \mathbb{R}$ is convex and ∇f is Lipschitz and $\int |\xi|^2 \mu(d\xi) < +\infty$, then

$$\mathcal{E}_\mu(f, n) = O(n^{-2/d}).$$

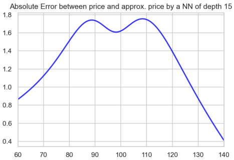
[a] G. Balazs, A. György, C. Szepesvari, Near-optimal max-affine estimators for convex regression,Proc. of the 18th International Conference on Artificial Intelligence and Statistics, PMLR 38:56-64, 2015

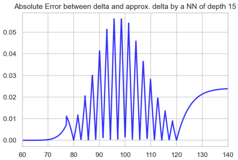
Differentiated network and AAD

- By *AAD* (Adjoint Automatic Differentiation) applied to the network, one can differentiate the (smoothenned) network with respect to its input x.
- The network provides a good approximation of the price function $x \mapsto \mathbb{E}\,\varphi(X_T^x)$, then the AAD-differentiated network approximates $\partial_x \mathbb{E}\, f(X_T^x)$ i.e. the δ-hedge at time 0.

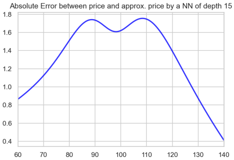
Preliminary tests

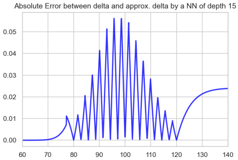
- Toy model: Call option with strike $K = 100$, $x \in [60, 140]$, $\sigma = 0.1$, $T = 1$.
- Dataset: size $N = 1000$ of $(x_i, Call(x_i))$ with $x_i \in [60, 140]$.
- Architecture: ICNN with n hidden layers and $d + 1 = 2$ units per layer.
- Starting state: $n = 15$ points $(x_1, \ldots, x_n)$ compute $\max_{k=1:15} \varphi_k(x)$ (closed form or Monte Carlo).



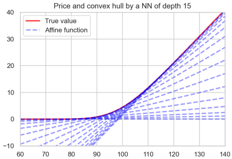


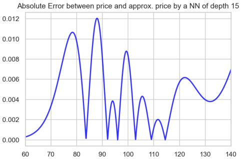
- Starting state:

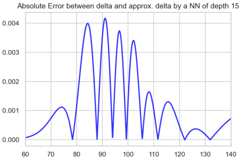




- After training by 50 epochs of Adam's method with step $\varepsilon = 10^{-3}$:







Next step: convex order...

Theorem (Convex order P., 2016)

If an $\mathbb{R}^d$-valued martingale diffusion satisfies

$$dY_t^x = x + \int_0^t \vartheta(Y_s^x)dW_s,$$

with W q-dim Brownian motion, and $x \mapsto \vartheta(x)$ is convex for the order $A \preccurlyeq B$ if $BB^* - AA^* \geq 0$ and $\vartheta \preccurlyeq \sigma$ then, for every $\mathbb{R}_+$-valued convex functional Φ :

$$\mathbb{E}\,\Phi(Y^{\vartheta,x}) \leq \mathbb{E}\,\Phi(Y^{\sigma,x}).$$

Merci de votre attention !