MLMC techniques for discontinuous functions

Mike Giles

(and many collaborators)

Mathematical Institute, University of Oxford

MCQMC 2022

July 19, 2022

Outline

- MLMC and the problem with discontinuous functions
- 7 approaches:
  - explicit smoothing
  - integration/differentiation
  - Malliavin calculus
  - conditional expectation
  - change-of-measure
  - splitting
  - adaptive sampling
- references

Several methods borrow ideas from computing sensitivities ("greeks") of the form

$$\frac{\partial}{\partial \alpha} \mathbb{E}\left[f(\omega, \alpha) \right]$$

(and one or two from improving smoothness for QMC integration)

Multilevel Monte Carlo

MLMC is based on the telescoping sum

$$\mathbb{E}[\widehat{P}_L] = \mathbb{E}[\widehat{P}_0] + \sum_{\ell=1}^{L} \mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$$

where $\widehat{P}_\ell$ represents an approximation to output P on level ℓ.

If $\widehat{Y}_\ell$ has expected value $\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$, with variance V_ℓ and cost C_ℓ, then for a given target RMS error ε, the number of independent samples on each level can be optimised to give overall cost

$$\varepsilon^{-2} \left(\sum_{\ell=0}^{L} \sqrt{C_\ell V_\ell} \right)^2 \sim \begin{cases} \varepsilon^{-2} C_0 V_0, & C_\ell V_\ell \to 0, \\ \varepsilon^{-2} L^2 C_L V_L, & C_\ell V_\ell \to \text{const}, \quad \ell \to \infty \\ \varepsilon^{-2} C_L V_L, & C_\ell V_\ell \to \infty. \end{cases}$$

Challenge 1 – nested expectation (VaR in finance)

Suppose we want to estimate $\mathbb{E}\left[\, f\left(\mathbb{E}[Z|X]\right)\right]$ where $\mathbb{E}[Z|X]$ has a bounded probability density function.

A simple MLMC treatment uses $M_\ell = 2^\ell M_0$ inner samples on level ℓ, so the cost C_ℓ is $O(M_\ell)$ and the estimator for a given outer sample X is

$$\widehat{Y}_\ell = f\left(\frac{1}{M_\ell}\sum_{m=1}^{M_\ell} Z^{(m)}\right) - f\left(\frac{1}{M_{\ell-1}}\sum_{m=1}^{M_{\ell-1}} Z^{(m+M_\ell)}\right)$$

with the $Z^{(m)}$ all generated independently conditional on X.

If $\mathbb{V}[Z|X]$ is finite and uniformly bounded, and f is Lipschitz, then $\widehat{Y}_\ell = O(M_\ell^{-1/2})$ so $V_\ell = O(M_\ell^{-1})$ and the complexity is $O(\varepsilon^{-2}|\log\varepsilon|^2)$.

However, for the Heaviside step function $H(\cdot)$, $\widehat{Y}_\ell = \pm 1$ with probability $O(M_\ell^{-1/2})$, so $V_\ell = O(M_\ell^{-1/2})$ and the complexity is $O(\varepsilon^{-5/2})$.

Challenge 2 – SDEs

In the case of an SDE

$$\mathrm{d}S_t = a(S_t)\,\mathrm{d}t + b(S_t)\,\mathrm{d}W_t$$

with an output quantity of interest $P \equiv f(S_T)$, the standard estimator is

$$\widehat{Y}_\ell = \Delta \widehat{P}_\ell \;\equiv\; \widehat{P}_\ell - \widehat{P}_{\ell-1}$$

where the same Brownian motion W_t is used for both $\widehat{P}_\ell$ and $\widehat{P}_{\ell-1}$, but with different uniform timesteps h_ℓ and $h_{\ell-1}$.

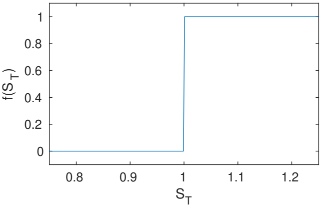
If f is Lipschitz, with constant L_f, then

$$V_\ell \;\leq\; \mathbb{E}\left[(\Delta \widehat{P}_\ell)^2\right] \;\leq\; L_f^2\, \mathbb{E}\left[(\widehat{S}_\ell - \widehat{S}_{\ell-1})^2\right]$$

so we have $V_\ell = O(h_\ell)$ for Euler-Maruyama discretisation, $V_\ell = O(h_\ell^2)$ for Milstein, and cost $C_\ell = O(h_\ell^{-1})$ in both cases.

Challenge 2 – SDEs (digital option in finance)

In mathematical finance, a digital call option payoff is 0 or 1, depending on whether S_T is below or above the strike K. The problem is that a small change in the path can give a big change in the payoff.



Using the Euler-Maruyama approximation the strong error is $O(h^{1/2})$,

$$\Longrightarrow \quad \widehat{S}_\ell - \widehat{S}_{\ell-1} = O(h_\ell^{1/2}).$$

An $O(h_\ell^{1/2})$ fraction of fine/coarse pairs straddle the strike,

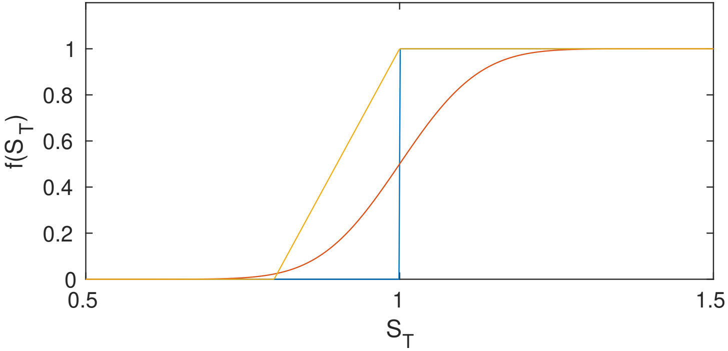
$$\Longrightarrow \quad V_\ell = O(h_\ell^{1/2}), \text{ and hence the complexity is } O(\varepsilon^{-5/2}).$$

Using the Milstein approximation the strong error is $O(h)$ so $V_\ell = O(h_\ell)$. This is better, but the kurtosis is $O(h_\ell^{-1})$ which causes problems in practice in estimating V_ℓ.

1) Explicit smoothing

Digital options are also a problem for IPA / pathwise sensitivity analysis, estimating the sensitivity of the expectation to a parameter change

One common solution to this is to explicitly smooth the payoff, and that can be used also for MLMC – involves a tradeoff between bias and variance



1) Explicit smoothing

G, Nagapetyan, Ritter (2015) use explicit smoothing for estimating CDFs.

For a scalar S_T, to estimate $C(x) = \mathbb{P}(S_T < x) = \mathbb{E}[H(x - S_T)]$ where $H(\cdot)$ is the Heaviside step function, the approach is

- use MLMC to estimate $C(x_j)$ for a set of spline points x_j
- interpolate with a cubic spline

To improve the MLMC variance, $H(x)$ is replaced by $H_\delta(x)$ which smooths H over a width of δ. Overall, have to balance four errors:

- SDE discretisation bias on finest level
- MLMC sampling error
- smoothing error
- interpolation error

Explicit smoothing also used recently by Xu, He, Wang (2020) for risk estimation

2) Integration/differentiation

Krumscheid, Nobile (2018) use a slightly different approach for estimating CDFs, based on

$$\frac{\mathrm{d}}{\mathrm{d}x}\, \mathbb{E}\left[\max(0, x - S_T)\right] = \mathbb{E}[\, H(x - S_T)\,]$$

- use MLMC to estimate $\mathbb{E}[\max(0, x_j - S_T)]$ for spline points x_j
- interpolate with a cubic spline
- differentiate to obtain an approximation to CDF $C(x)$

This avoids the smoothing error, but differentiating the cubic spline amplifies the noise in the spline data.

3) Malliavin calculus

On a similar note, Altmayer & Neuenkirch (2015) used Malliavin calculus integration by parts to handle discontinuous payoffs with the Heston model

Used on its own it improves the asymptotic behaviour, but makes the variance on coarse levels worse.

To address this, they split the payoff into a smooth part (handled by standard MLMC) and a compact-support discontinuous part (handled using Malliavin MLMC)

Again one key lesson is that techniques which help with computing sensitivities can also help with MLMC.

4) Conditional expectation

For the Milstein discretisation, one "fix" for digital options is to use E-M approximation for the final timestep, then take conditional expectation over final fine path Brownian increment ΔW_N.

For fine path with timestep h_ℓ,

$$\widehat{S}^f_T = \widehat{S}^f_{T-h_\ell} + a(\widehat{S}^f_{T-h_\ell})\, h_\ell + b(\widehat{S}^f_{T-h_\ell})\, \Delta W_N,$$

$$\Longrightarrow \quad \widehat{P}^f_\ell = \Phi\left(\frac{\widehat{S}^f_{T-h_\ell} + a(\widehat{S}^f_{T-h_\ell})\, h_\ell - K}{b(\widehat{S}^f_{T-h_\ell})\, \sqrt{h_\ell}} \right)$$

while for the coarse path with $h_{\ell-1} = 2\, h_\ell$,

$$\widehat{S}^c_T = \widehat{S}^c_{T-h_{\ell-1}} + a(\widehat{S}^c_{T-h_{\ell-1}})\, h_{\ell-1} + b(\widehat{S}^c_{T-h_{\ell-1}})\, (\Delta W_{N-1} + \Delta W_N),$$

$$\Longrightarrow \quad \widehat{P}^c_{\ell-1} = \Phi\left(\frac{\widehat{S}^c_{T-h_{\ell-1}} + a(\widehat{S}^c_{T-h_{\ell-1}})\, h_\ell + b(\widehat{S}^c_{T-h_{\ell-1}})\, \Delta W_{N-1} - K}{b(\widehat{S}^c_{T-h_{\ell-1}})\, \sqrt{h_\ell}} \right)$$

4) Conditional expectation

Analysis (G, Debrabant, Roessler, 2019) proves $V_\ell \approx O(h_\ell^{3/2})$ and the kurtosis is $O(h_\ell^{-1/2})$, so much better.

Heuristically, this is because there is an $O(h_\ell^{1/2})$ probability of paths being within $O(h_\ell^{1/2})$ of the strike K, and for these

$$\widehat{S}_\ell - \widehat{S}_{\ell-1} = O(h_\ell), \quad \frac{\partial \widehat{P}}{\partial S_T} = O(h_\ell^{-1/2}), \quad \Longrightarrow \quad \Delta \widehat{P}_\ell = O(h_\ell^{1/2})$$

Unfortunately, the conditional expectation approach does not help with the E-M discretisation where

$$\widehat{S}_\ell - \widehat{S}_{\ell-1} = O(h_\ell^{1/2}) \quad \Longrightarrow \quad \Delta \widehat{P}_\ell = O(1)$$

4) Conditional expectation

Conditional expectation is a standard technique for smoothing the payoff to enable IPA/pathwise sensitivity calculations (L'Ecuyer, Glasserman)

Another example is a down-and-out barrier option, where the option is knocked out if the path drops below a certain value.

Payoff can be smoothed by computing probability of this happening, conditional on computed path approximations at discrete timesteps.

Again, this works well for both pathwise sensitivity analysis and MLMC.

(G, 2008, Burgos, G, 2012)

4) Conditional expectation

Achtsis, Cools, Nuyens (2013), Bayer, Siebenmorgen, Tempone (2018), Bayer, Ben Hammouda, Tempone (2020) split the random inputs W into a scalar Z and the remainder W_r.

They then express the desired MLMC level ℓ expectation as

$$\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}] = \mathbb{E}\left[\ \mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1} \,|\, W_r]\ \right]$$

and observe that in many financial applications it is possible to perform this split in a way such that

$$\mathbb{E}[\widehat{P}_\ell \,|\, W_r], \quad \mathbb{E}[\widehat{P}_{\ell-1} \,|\, W_r]$$

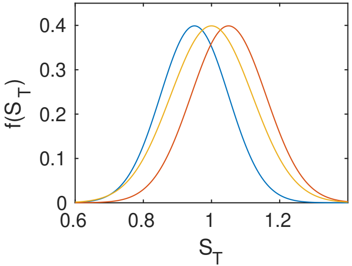
are smooth functions of W_r, and can be evaluated analytically or very accurately by root-finding in Z to locate the one discontinuity.

For a scalar SDE, Z could be the terminal value of the driving Brownian motion, and W_r would be the other random variables required for a Brownian Bridge construction of the Brownian increments.

5) Change of measure

Another approach with the Milstein discretisation is to use a change of measure – similar to LRM for sensitivity analysis

For both the fine and coarse paths, we have conditional Gaussian distributions for $\widehat{S}_T$, with slightly different means and variances.



Can perform a change of measure to the same Gaussian distribution, and then pick the same sample for both paths, giving

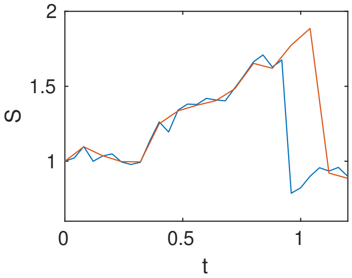
$$\widehat{P}_\ell - \widehat{P}_{\ell-1} = \widehat{P}(\widehat{S}_T)\,(R_\ell - R_{\ell-1}) = O(h_\ell^{1/2})$$

where $R_\ell, R_{\ell-1}$ are the respective Radon-Nikodym derivatives – works well in multiple dimensions where often cannot evaluate the analytic conditional expectation (Burgos, 2014)

Doesn't help with EM discretisation because $R_\ell - R_{\ell-1} = O(1)$

5) Change of measure

An earlier example of its use was for a Merton-style jump-diffusion SDE with path-dependent jump rate $\lambda(S,t)$ (Xia, G, 2012, Xia, 2014)



Problem is that coarse and fine paths will jump at different times; one might jump just before T, and the other just after

Solution: use Glasserman & Merener thinning technique, over-sampling possible jump times using rate $\lambda_{sup} > \lambda(S,t)$, and combine with change of measure for identical acceptance/rejection decision for fine/coarse paths.

Leads to an estimator which looks like

$$\widehat{P}_\ell \, R_\ell - \widehat{P}_{\ell-1} \, R_{\ell-1}$$

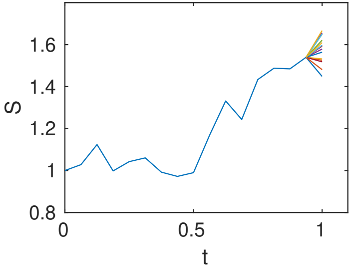
and gives $V_\ell = O(h_\ell^2)$ when combined with Milstein discretisation.

6) Splitting

Back again to the digital option.

Conditional expectation can be estimated by averaging over a number of independent samples for the final Brownian increment.

$O(h_\ell^{-1})$ final samples can be used without increasing the path cost significantly.



This is sufficient to reduce V_ℓ to about the same level as using the analytic conditional expectation.

Bonus: can use more accurate Milstein method for final timestep.

Burgos, G (2012), Burgos (2014) also used splitting for MLMC for pathwise sensitivity analysis for put/call options.

6) Splitting

Bernal, G (2019) used splitting for Feynman-Kac functionals arising for stopped diffusions – SDE calculations which terminate when the path leaves the domain.

The issue here is that when a fine path exits, there is an $O(h_\ell^{1/2})$ probability that the corresponding coarse path does not leave until much later.

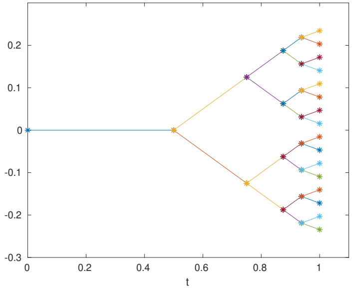
This is solved by estimating a conditional expectation by splitting the coarse path into $O(h_\ell^{-1/2})$ independent sub-simulations.

V_ℓ is improved from $O(h_\ell^{1/2})$ to approximately $O(h_\ell)$.

6) Splitting

Digital options using the E-M discretisation can be handled efficiently by repeated path splitting.

When simulating on a unit time interval, split into 2 paths at $t = 1/2$, then 4 paths at $t = 3/4$, then 8 paths at $t = 7/8$, and so on, with final split when there is just one coarse timestep left.



Each path continuation is independent of all of the others, i.e. uses its own Brownian motion. Computational cost per sample is $O(h_\ell^{-1} |\log h_\ell|)$.

This effectively smoothes out the discontinuity, and the variance is reduced to approximately $O(h_\ell)$ as with Lipschitz payoffs.

Paper with Haji-Ali in preparation.

6) Splitting

Pros and cons of repeated splitting:

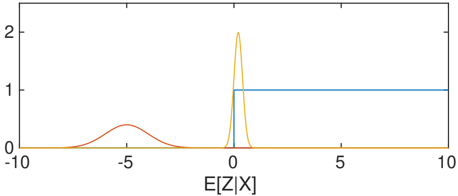
- fairly easy to implement
- near-optimal complexity for digital options with EM discretisation, and with a moderate kurtosis
- non-adaptive so can handle multiple discontinuities at same time (e.g. for CDF estimation)
- restricted to discontinuous functions of terminal state? (no obvious way to handle barrier options)
- maybe generalisable to SPDEs

7) Adaptive sampling

We return to the nested expectation problem: $\mathbb{E}\left[\, H\left(\mathbb{E}[Z|X]\right) \right]$ and note that only need an accurate estimate of $\mathbb{E}[Z|X]$ when it is near zero.

Building on adaptive sampling within MC (Broadie, Du, Moallemi, 2011), adaptive sampling combined with MLMC (G., Haji-Ali, 2019) uses

- $M_\ell = 2^\ell M_0$ inner samples when $|\mathbb{E}[Z|X]| \gg \sqrt{\mathbb{V}[Z|X]/(2^\ell M_0)}$
- $M_\ell = 4^\ell M_0$ inner samples when $|\mathbb{E}[Z|X]| = O(\sqrt{\mathbb{V}[Z|X]/(4^\ell M_0)})$
- $2^\ell M_0 < M_\ell < 4^\ell M_0$ for intermediate values



Leads to $C_\ell \sim 2^\ell$, $V_\ell \sim 2^{-\ell}$ and hence complexity of roughly $O(\varepsilon^{-2})$

7) Adaptive sampling

Haji-Ali, Spence, Teckentrup (2021) extend this to

$$\mathbb{P}[G \in \Omega] \equiv \mathbb{E}[1_{G\in\Omega}]$$

where G is a d-dimensional random variable which cannot be sampled directly – the two examples they consider are precisely the two challenges in this talk.

Adaptive sampling for the digital option with EM discretisation uses

- $h_\ell = 2^{-\ell}$ when $|\widehat{S}_\ell - K|$ large
- $h_\ell = 4^{-\ell}$ when $|\widehat{S}_\ell - K|$ small
- $2^{-\ell} < h_\ell < 4^{-\ell}$ for intermediate values

A Brownian bridge construction is used when the timestep needs to be refined from its base level.

Again leads to $C_\ell \sim 2^\ell$, $V_\ell \sim 2^{-\ell}$ and hence complexity of roughly $O(\varepsilon^{-2})$

7) Adaptive sampling

In earlier research, Elfverson, Hellman, Malqvist (2016) considered estimation of $\mathbb{E}[H(X)]$ where X cannot be sampled exactly but there is a sequence of approximations $X'_0, X'_1, X'_2, \ldots X$ of increasing accuracy and increasing cost.

Motivated by PDE applications with a well-behaved truncation error so that there are uniform geometric bounds on $|X'_j - X|$, level ℓ uses

$$\widehat{X}_\ell = X'_j, \quad j = \min\{\ell, \min j : |\widehat{X}'_j - X| < |X|\}$$

and achieves similarly good MLMC benefits.

The idea is essentially the same as in the work of Haji-Ali *et al* but requiring a uniform bound on $|X'_j - X|$ is significantly more restrictive than needing bounds on $\mathbb{E}[\,|X'_j - X|^q]$ for some $q > 2$.

7) Adaptive sampling

Pros and cons of adaptive sampling:

- also fairly simple to implement
- near optimal complexity, but often with a bad kurtosis
- adaptation is specific to output function – for CDF estimation would have to adapt for each spline point
- analysis of Haji-Ali, Spence, Teckentrup can be generalised to product of indicator function and Lipschitz function, $\mathbb{E}[1_{G\in\Omega} f(S)]$, and so can handle barrier options

Conclusions

- most MLMC applications use "plain" MLMC with no need for any of these techniques – good strong convergence implies small variance
- for applications with discontinuities, there is a growing toolkit of techniques to consider, with the same techniques being used in widely differing applications.
- in many cases, ideas have been adapted from sensitivity analysis which also has problems with discontinuous functionals

3 talks later this week:

- Chiheb Ben Hammouda on conditional expectation (numerical smoothing) on Wednesday at 14:00
- Al Haji-Ali on repeated splitting (branching) on Thursday at 17:00
- Jonny Spence on adaptive sampling on Friday at 9:00

References

N. Achtsis, R. Cools, D.Nuyens. 'Conditional sampling for barrier option pricing under the LT method',SIAM Journal on Financial Mathematics, 4(1):327-352, 2013.

M. Altmayer, A. Neuenkirch. 'Multilevel Monte Carlo quadrature of discontinuous payoffs in the generalized Heston model using Malliavin integration by parts'. SIAM Journal on Financial Mathematics, 6(1):22-52, 2015.

C. Bayer, C. Ben Hammouda, R. Tempone. 'Multilevel Monte Carlo combined with numerical smoothing for robust and efficient option pricing and density estimation'. arXiv preprint 2003.05708, 2020.

C. Bayer, M. Siebenmorgen, R. Tempone. 'Smoothing the payoff for efficient computation of basket option prices'. Quantitative Finance, 18(3):491-505, 2018.

M. Broadie, Y. Du, C.C. Moallemi, 'Efficient risk estimation via nested sequential simulation', Management Science, 57:1172-1194, 2011.

S. Burgos, G. 'Computing Greeks using multilevel path simulation'. pp.281-296 in Monte Carlo and Quasi-Monte Carlo Methods 2010, Springer, 2012.

S. Burgos. 'The computation of Greeks with multilevel Monte Carlo'. PhD thesis, University of Oxford, 2014.

References

D. Elfverson, F. Hellman, A. Malqvist. 'A multilevel Monte Carlo method for computing failure probabilities'. SIAM/ASA Journal on Uncertainty Quantification, 4(1):312-330, 2016.

G. 'Improved multilevel Monte Carlo convergence using the Milstein scheme'. pp.343-358, in Monte Carlo and Quasi-Monte Carlo Methods 2006, Springer, 2008.

G. 'Multilevel Monte Carlo methods'. Acta Numerica, 24:259-328, CUP, 2015.

G, F. Bernal. 'Multilevel estimation of expected exit times and other functionals of stopped diffusions'. SIAM/ASA Journal on Uncertainty Quantification, 6(4):1454-1474, 2018.

G, K. Debrabant, A. Roessler. 'Analysis of multilevel Monte Carlo path simulation using the Milstein discretisation'. Discrete and Continuous Dynamical Systems – series B, 24(8):3881-3903, 2019.

G, A.-L. Haji-Ali. 'Multilevel nested simulation for efficient risk estimation'. SIAM/ASA Journal on Uncertainty Quantification, 7(2):497-525, 2019.

G, T. Nagapetyan, K. Ritter. 'Multi-Level Monte Carlo approximation of distribution functions and densities'. SIAM/ASA Journal on Uncertainty Quantification, 3:267-295, 2015.

References

A.-L. Haji-Ali, J. Spence, A. Teckentrup. 'Adaptive multilevel Monte Carlo for probabilities', SIAM Journal on Numerical Analysis, to appear, 2022.
(also arXiv preprint 2107.09148)

S. Krumscheid, F. Nobile. 'Multilevel Monte Carlo approximation of functions'. SIAM/ASA Journal on Uncertainty Quantification, 6(3):1256-1293, 2018.

Y. Xia, G. 'Multilevel path simulation for jump-diffusion SDEs', pp.695-708 in Monte Carlo and Quasi-Monte Carlo Methods 2010, Springer, 2012.

Y. Xia. 'Multilevel Monte Carlo for jump processes'. PhD thesis, University of Oxford, 2014.

Z. Xu, Z. He, X. Wang. 'Efficient risk estimation via nested multilevel quasi-Monte Carlo simulation'. arXiv preprint 2011.11898, 2020.

Webpages:

`http://people.maths.ox.ac.uk/gilesm/mlmc/`

`http://people.maths.ox.ac.uk/gilesm/mlmc.html`

`http://people.maths.ox.ac.uk/gilesm/mlmc_community.html`