Supplementary Material for paper Adaptive Stratified Sampling for Monte-Carlo integration of Differentiable functions

A Numerical Experiments

We provide some experiments illustrating how LMC-UCB works, and compare its efficiency to that of crude Monte-Carlo and Uniform stratified Monte-Carlo.

We first illustrate on an example, in Figure 2, the sampling scheme. We have launched LMC-UCB on the function displayed in Figure 2 (i.e. $f(x) = \sin(1/(x+0.1)) + \mathbb{I}\{x > 0.9\}\sin(1/(x-0.7))$). We chose this function since its variations are quite heterogeneous in the domain $[0, 1]$. We considered a budget of $n = 100$, and took as parameter $A = 10$. K_n and $\bar{S}$ are defined as in Figure 1.

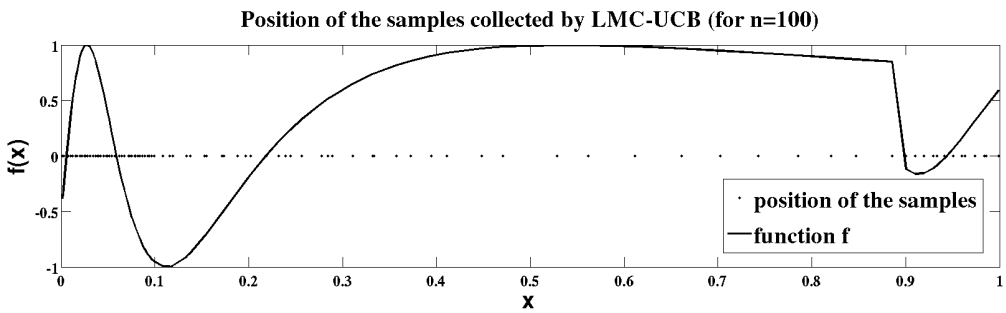

Figure 2: Position of the samples collected by LMC-UCB.

We observe that, as expected, the algorithm allocates more points in parts of the domain where the function has larger variations and, additional to that, it spreads the points on the domain so that every region is covered (in a similar spirit to what low-discrepancy schemes would do).

We also compare, for this function, the mean squared error of crude Monte-Carlo, uniform stratified Monte-Carlo and LMC-UCB, for different values of n. We average the mean squared error of the estimate returned by each method on 10000 runs. We have the following performances for each method (displayed in Figures 3 and 4).

As expected, the mean square error decreases faster than $1/n$ for uniform stratified Monte-Carlo and LMC-UCB. These methods are also more efficient than crude Monte-Carlo (up to 100 times more efficient on this function), which makes sense since the function that we integrate is differentiable (and then the rate for LMC-UCB and Uniform stratified Monte-Carlo is of order $O(n^{-1-2/d})$). The gain in efficiency when compared to crude Monte-Carlo however decreases with the dimension, as explained in Subsection 5.3. We observe that LMC-UCB is more efficient than uniform stratified Monte-Carlo, which is a minimax-optimal strategy in the class of non-adaptive strategies.

B Poof of Lemma 1

Step 0: Decomposition of the variance Let $\Omega = (\Omega_k^n)_{0<n<+\infty, k\leq n}$ be a sequence of partitions of $[0, 1]^d$ in n hyper-cubic strata such that the maximum diameter of the strata in the partitions converges to 0 when n goes to infinity. In each of those strata, there is a point.

Let n be the number of points, and $k \leq n$ be an index. Let $a_{n,k}$ be a point of the stratum Ω_k^n. Let us assume that f is differentiable, that it's derivative ∇f is continuous, and let us also assume that

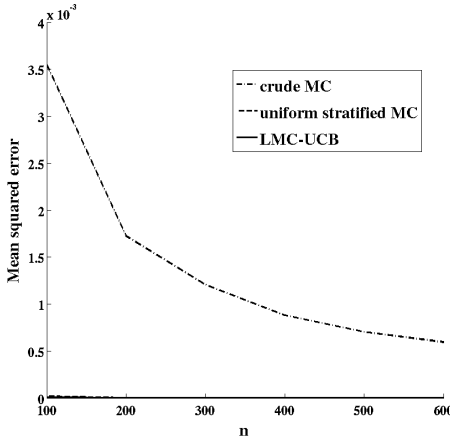

Figure 3: Mean squared error w.r.t. the integral of f of crude Monte-Carlo, uniform stratified Monte-Carlo and LMC-UCB, in function of the budget n. Since crude Monte-Carlo is approximately 100 times less efficient than the two other strategies, their curves are shrinked and not very visible.

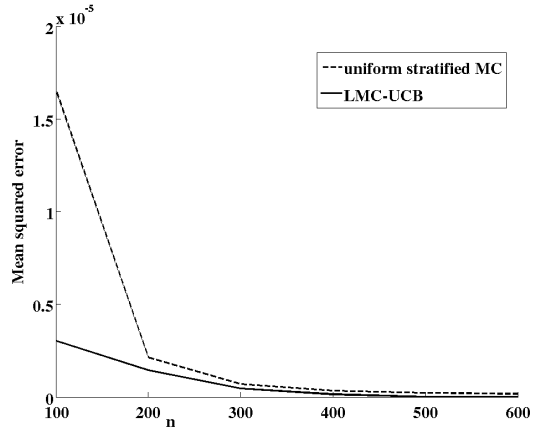

Figure 4: Zoom on the mean squared error w.r.t. the integral of f of uniform stratified Monte-Carlo and LMC-UCB, in function of the budget n.

$||\nabla f(u)||_2^2 = \sum_{i=1}^d \left(\frac{\partial f(u)}{\partial x_i}\right)^2$ is such that $\int ||\nabla f(x)||_2^2 dx$ is bounded. In that case, $\forall x \in \Omega_k^n$, there exists $u_{n,k,x} \in \Omega_k^n$ such that we have $f(x) - f(a_k) = \langle \nabla f(u_{n,k,x}), x - a_{n,k}\rangle$ (intermediate values theorem). Note also that we have in that case $\mu_{n,k} = f(a_{n,k}) + \frac{1}{w_{n,k}} \int_{\Omega_k^n} \langle \nabla f(u_{n,k,x}), x - a_{n,k}\rangle dx$ where $a_{n,k}$ is the center of the stratum Ω_k^n. We thus have:

$$
\begin{aligned}
\sigma_{n,k}^2 =& \frac{1}{w_{n,k}} \int_{\Omega_k^n} (f(x) - f(a_{n,k}))^2 dx \\
=& \frac{1}{w_{n,k}} \int_{\Omega_k^n} \Big(\langle \nabla f(u_{n,k,x}), x - a_{n,k}\rangle - \frac{1}{w_{n,k}} \int_{\Omega_k^n} \langle \nabla f(u_{n,k,y}), y - a_{n,k}\rangle dy \Big)^2 dx \\
=& \frac{1}{w_{n,k}} \int_{\Omega_k^n} \Big(\langle \nabla f(u_{n,k,x}), x - a_{n,k}\rangle \Big)^2 dx - \Big(\frac{1}{w_{n,k}} \int_{\Omega_k^n} \langle \nabla f(u_{n,k,y}), y - a_{n,k}\rangle dy \Big)^2 \\
=& \frac{1}{w_{n,k}} \int_{[0,1]^d} \Big(\langle \nabla f(u_{n,k,x}) \mathbb{I}\left\{\Omega_k\right\}, (x - a_{n,k}) \mathbb{I}\left\{\Omega_k^n\right\} \rangle \Big)^2 dx \\
& - \Big(\frac{1}{w_{n,k}} \int_{[0,1]^d} \langle \nabla f(u_{n,k,y}) \mathbb{I}\left\{\Omega_k^n\right\}, (y - a_{n,k}) \mathbb{I}\left\{\Omega_k^n\right\} \rangle dy \Big)^2 .
\end{aligned}
$$

Step 1: Convergence of σ_k when the size of the strata goes to 0 Let $x \in [0,1]^d$. Note that as as $(\Omega_k^n)_{k \le n}$ is a partition, there is a $k_{n,x}$ such that $x \in \Omega_{k_{n,x}}^n$.

Note first that ∇f is continuous. This means that $\forall \epsilon, \exists \eta / \forall y \in \mathcal{B}_2(x, \eta), ||\nabla f(y) - \nabla f(x)||_2 \le \epsilon$. Let $\epsilon > 0$ and n sufficiently large (any n larger than some given horizon n'), the maximum diameter of $\Omega_{k_{n,x}}^n$ is smaller than η. Let $y \in \Omega_{k_{n,x}}^n$. As $u_{n,k_{n,x},y} \in \Omega_{k_{n,x}}^n$, we know that $||u_{n,k_{n,x},y} - x|| \le \eta$ and that we thus have $||\nabla f(u_{n,k_{n,x},y}) - \nabla f(x)||_2 \le \epsilon$. This means that $\nabla f(u_{n,k_{n,x},y})$ converges point-wise to $\nabla f(x)$.

Note also that we have by Cauchy-Schwartz that

$$
\begin{aligned}
\frac{1}{w_{n,k_{n,x}}^{2/d}} \Big(\langle \nabla f(u_{n,k_{n,x},y}), (y - a_{n,k_{n,x}}) \rangle \Big)^2 \mathbb{I}\left\{\Omega_{k_{n,x}}^n\right\} \le& \frac{1}{w_{n,k_{n,x}}^{2/d}} ||\nabla f(u_{n',k_{n',x},y})||_2^2 ||y - a_{n,k_{n,x}}||_2^2 \mathbb{I}\left\{\Omega_{k_{n,x}}^n\right\} \\
\le& d||\nabla f(u_{n,k_{n,x},y})||_2^2 \le dL^2 .
\end{aligned}
$$

As $\nabla f(u_{n,k_{n,x},y})$ converges point-wise with n to $\nabla f(x)$, and as $\frac{1}{w_{n,k_{n,x}}^{2/d}}\Big(\langle \nabla f(u_{n,k_{n,x},y}),(y-a_{n,k_{n,x}})\rangle\Big)^2 \leq dL^2$, we have by the Theorem of Dominated convergence, that

$$\begin{aligned}
&\lim_{n\to+\infty}\frac{1}{w_{n,k_{n,x}}^{1+2/d}}\int_{[0,1]^d}\Big(\langle \nabla f(u_{n,k_{n,x},y}),(y-a_{n,k_{n,x}})\rangle\Big)^2\mathbb{I}\left\{\Omega_{k_{n,x}}^n\right\}dy\\
&\lim_{n\to+\infty}\frac{1}{w_{n,k_{n,x}}^{1+2/d}}\int_{[0,1]^d}\Big(\langle \lim_{n\to+\infty}\nabla f(u_{n,k_{n,x},y}),(y-a_{n,k_{n,x}})\rangle\Big)^2\mathbb{I}\left\{\Omega_{k_{n,x}}^n\right\}dy\\
&\lim_{n\to+\infty}\frac{1}{w_{n,k_{n,x}}^{1+2/d}}\int_{[0,1]^d}\Big(\langle \nabla f(x),(y-a_{n,k_{n,x}})\rangle\Big)^2\mathbb{I}\left\{\Omega_{k_{n,x}}^n\right\}dy\\
&=\lim_{n\to+\infty}\frac{1}{w_{n,k_{n,x}}^{1+2/d}}\frac{||\nabla f(x)||_2^2 w_{n,k_{n,x}}^{1+2/d}}{12}\\
&=\frac{||\nabla f(x)||_2^2}{12}.
\end{aligned}$$

In the same way, we have that

$$\begin{aligned}
&\lim_{n\to+\infty}\frac{1}{w_{n,k_{n,x}}^{1+2/d}}\Big(\int_{[0,1]^d}\Big(\langle \nabla f(u_{n,k_{n,x},y}),(y-a_{n,k_{n,x}})\rangle\mathbb{I}\left\{\Omega_{k_{n,x}}^n\right\}dy\Big)^2\\
&\lim_{n\to+\infty}\frac{1}{w_{n,k_{n,x}}^{1+2/d}}\Big(\int_{[0,1]^d}\langle \lim_{n\to+\infty}\nabla f(u_{n,k_{n,x},y}),(y-a_{n,k_{n,x}})\rangle\mathbb{I}\left\{\Omega_{k_{n,x}}^n\right\}dy\Big)^2\\
&\lim_{n\to+\infty}\frac{1}{w_{n,k_{n,x}}^{1+2/d}}\Big(\int_{[0,1]^d}\langle \nabla f(x),(y-a_{n,k_{n,x}})\rangle\mathbb{I}\left\{\Omega_{k_{n,x}}^n\right\}dy\Big)^2\\
&=\lim_{n\to+\infty}\frac{1}{w_{n,k_{n,x}}^{1+2/d}}w_{n,k_{n,x}}^{1+2/d}\Big(a_{n,k_{n,x}}-a_{n,k_{n,x}}\Big)\\
&=0.
\end{aligned}$$

Let us call $g_{n,\Omega}(x)=\sum_{k=1}^n\frac{\sigma_{n,k}^2}{w_{n,k}^{1/2d}}\mathbb{I}\left\{\Omega_k^n\right\}(x)=\frac{\sigma_{n,k_{n,x}}^2}{w_{n,k_{n,x}}^{1/2d}}$. The last two inequalities prove, $\forall x$, point-wise convergence of $g_{n,\Omega}(x)$ to $\frac{||\nabla f(x)||_2^2}{12}$:

Step 2: Optimal allocation and minimum for the asymptotic variance There is one point pulled at random per stratum. The variance of the estimate given by such an allocation is

$$\sum_{k=1}^n w_{n,k}^2\sigma_{n,k}^2=\sum_{k=1}^n w_{n,k}\times w_{n,k}^{1+2/d}\times\frac{\sigma_{n,k}^2}{w_{n,k}^{2/d}}.$$

Define $s_{n,\Omega}(x)=\sum_{k=1}^n\frac{1}{nw_{n,k}}\mathbb{I}\left\{\Omega_k^n\right\}(x)$. Note first that

$$1=\frac{1}{n}\sum_{k=1}^n 1=\int_{[0,1]^d}s_{n,\Omega}(x)dx,$$

and that

$$s_{n,\Omega}(x)>0.$$

One has also for the variance of the estimate that

$$\sum_{k=1}^n w_{n,k}^2\sigma_{n,k}^2=\frac{1}{n^{1+2/d}}\int_{[0,1]^d}g_{n,\Omega}(x)\frac{1}{s_{n,\Omega}(x)^{1+2/d}}dx.$$

By using the result of the previous step, one has (for every sequence Ω where the diameter of the strata converge uniformly to 0), point-wise convergence of $g_{n,\Omega}(x)$ to $\frac{||\nabla f(x)||_2^2}{12}$ when n goes to infinity.
This leads to, by using Fatou's Lemma

$$\begin{aligned}
&\lim_{n\to+\infty}\inf \int_{[0,1]^d} g_{n,\Omega}(x)\frac{1}{s_{n,\Omega}(x)^{1+2/d}}dx \\
&\geq \int_{[0,1]^d} \lim_{n\to+\infty}\inf \Big(g_{n,\Omega}(x)\frac{1}{s_{n,\Omega}(x)^{1+2/d}}\Big)dx \\
&\geq \int_{[0,1]^d} \inf_{s:s\geq 0,\int s=1} \frac{||\nabla f(x)||_2^2}{12}\frac{1}{s(x)^{1+2/d}}dx.
\end{aligned}$$

One thus wants then to find the function $s(x)$ that minimizes this limit. One thus wants to solve in each point x the program $\inf_s \frac{||\nabla f(x)||_2^2}{12}\frac{1}{s(x)^{1+2/d}}$ such that $s\geq 0$ and $\int_{[0,1]^d} s(x)dx = 1$.
The solution (by just writing Lagragian) is

$$s^*(x) = \frac{(||\nabla f(x)||_2)^{\frac{d}{d+1}}}{\int_{[0,1]^d}(||\nabla f(u)||_2)^{\frac{d}{d+1}}du}.$$

By plugging it in the bound, one obtains

$$\begin{aligned}
&\lim_{n\to+\infty}\inf \int_{[0,1]^d} g_{n,\Omega}(x)\frac{1}{s_{n,\Omega}(x)^{1+2/d}}dx \\
&\geq \frac{\Big(\int_{[0,1]^d}(||\nabla f(x)||_2)^{\frac{d}{d+1}}dx\Big)^{2\frac{(d+1)}{d}}}{12}.
\end{aligned}$$

Note that the previous result holds for any sequence of partitions $(\Omega_n)_n$ where the diameter of each stratum converges uniformly to 0. One finally has, using that, that the minimum possible asymptotic variance is bounded by

$$\lim_{n\to+\infty}\inf_{\Omega} n^{1+2/d}\sum_{k=1}^{n} w_{n,k}^2\sigma_{n,k}^2 \geq \frac{\Big(\int_{[0,1]^d}(||\nabla f(x)||_2)^{\frac{d}{d+1}}dx\Big)^{2\frac{(d+1)}{d}}}{12},$$

and we thus obtain the desired result.

C Proof of Lemmas 3

Upper bound on the standard deviation: The upper confidence bounds $B_{k,t}$ used in the MC-UCB algorithm is an elaboration in the specific case of Lipschitz function on Theorem 10 in [8] (a variant of this result is also reported in [1]). We state here a main Lemma.

Lemma 4 *Assume that the function f from which the data is collected is differentiable, and that $||\nabla f(x)||_2$ is bounded by L, and $n\geq 2$. Define the following event*

$$\xi = \xi_{K,n}(\delta) = \bigcap_{1\leq k\leq K,}\left\{\left|\sqrt{\frac{1}{\bar{S}-1}\sum_{i=1}^{\bar{S}}\Big(X_{k,i}-\frac{1}{\bar{S}}\sum_{j=1}^{\bar{S}}X_{k,j}\Big)^2}-\sigma_k\right|\leq 2L\sqrt{d}(\frac{w_k}{\bar{S}})^{1/d}\sqrt{\frac{\log(2K/\delta)}{\bar{S}}}\right\}. \tag{9}$$

The probability of ξ is bounded by $1-\delta$.

Note that the first term in the absolute value in Equation 9 is the empirical standard deviation of arm k computed as in Equation 8 for t samples. The event ξ plays an important role in the proofs of this section and a number of statements will be proved on this event.

We now provide the proof of Lemma 4.

Let us assume that f is such that $||\nabla f||_2 \leq L$. Let us consider a small box Ω_w of size w and such that $\Omega_w = \prod_{i=1}^d [a_i - \frac{w^{1/d}}{2}, a_i + \frac{w^{1/d}}{2}]$. As $||\nabla f||_2 \leq L$, we know that $|f(x) - \frac{1}{w}\int_{\Omega_w} f(u)du| \leq L\sqrt{d}w^{1/d}$.
If U is a random variable on Ω_w and $X = f(U)$, then

$$|X - \mu| \leq L\sqrt{d}w^{1/d},$$

where $\mu = \frac{1}{w}\int_{\Omega_w} f(u)du$.

Note first that for algorithm LMC-UCB, the $\bar{S}$ first samples are each sampled in an hypercube of measure $\frac{w_k}{\bar{S}}$, and all of those hypercubes form a partition of the domain.
Using a large deviation bound on the variance, e.g. the one in [8], we can deduce that with probability $1 - 2\delta$

$$|\sqrt{\frac{1}{\bar{S}-1}\sum_{i=1}^{\bar{S}}\Big(X_{k,i} - \frac{1}{\bar{S}}\sum_{j=1}^{\bar{S}} X_{k,j}\Big)^2} - \sigma_k| \leq b\sqrt{\frac{2\log(1/\delta)}{\bar{S}-1}},$$

where b is a bound on the random variables $X_i - \mu_i$. One gets because $|X_{k,i} - \mu_{k,i}| \leq \sqrt{d}L(\frac{w_k}{t})^{1/d}$ (where $\mu_{k,i}$ is the mean of the function on the hypercube where point $X_{k,i}$ is sampled and because $t \geq 2$

$$|\sqrt{\frac{1}{\bar{S}-1}\sum_{i=1}^{\bar{S}}\Big(X_{k,i} - \frac{1}{\bar{S}}\sum_{j=1}^{\bar{S}} X_{k,j}\Big)^2} - \sigma_k| \leq 2L\sqrt{d}(\frac{w_k}{\bar{S}})^{1/d}\sqrt{\frac{\log(1/\delta)}{\bar{S}}}.$$

Then by doing a simple union bound on (k, t), we obtain the result.

The following Corollary holds.

Corollary 1 *On the event ξ, $\forall k \leq K$,*

$$|\hat{\sigma}_{k,K\bar{S}} - \sigma_k| \leq 2L\sqrt{d}\sqrt{\log(2K/\delta)}\frac{w_k^{1/d}}{\bar{S}^{\frac{d+2}{2d}}}$$

By concavity, we also have the following Corollary.

Corollary 2 *On the event ξ, there is $\forall k \leq K$ that*

$$|\hat{\sigma}_{k,K\bar{S}}^{\frac{d}{d+1}} - \sigma_k^{\frac{d}{d+1}}| \leq A\frac{w_k^{\frac{1}{d+1}}}{\bar{S}^{\frac{d+2}{2(d+1)}}},$$

where $A = (2L\sqrt{d}\sqrt{\log(2K/\delta)})^{\frac{d}{d+1}}$.

The number of sub-strata Let k be an index. Let us call $C_k = \frac{w_k^{\frac{d}{d+1}}\left(\hat{\sigma}_{k,K\bar{S}} + A(\frac{w_k}{\bar{S}})^{1/d}\sqrt{\frac{1}{\bar{S}}}\right)^{\frac{d}{d+1}}}{\sum_{i=1}^K w_i^{\frac{d}{d+1}}\left(\hat{\sigma}_{i,K\bar{S}} + A(\frac{w_i}{\bar{S}})^{1/d}\sqrt{\frac{1}{\bar{S}}}\right)^{\frac{d}{d+1}}}(n - K\bar{S})$.

Stratum Ω_k is subdivided in $S_k = \max\left[\bar{S}, \lfloor C_k^{1/d}\rfloor^d\right]$ substrata, composing the sub-partition $\mathcal{N}_k$.

Note first that $\sum_{k=1}^K S_k \leq n$ as $\sum_{k=1}^K C_k = n - K\bar{S}$. As the samples are always picked in sub-strata that have the less points, it ensures that there is at least one point per sub-stratum.

On ξ, we have because of Corollary 2 that

$$\begin{aligned}
C_k &\geq \frac{w_k^{\frac{d}{d+1}}\sigma_k^{\frac{d}{d+1}}}{\sum_{i=1}^K w_i^{\frac{d}{d+1}}\Big(\sigma_i^{\frac{d}{d+1}} + 2A\frac{w_i^{\frac{1}{d+1}}}{\bar{S}^{\frac{d+2}{2(d+1)}}}\Big)}(n-K\bar{S}) \\
&\geq \frac{w_k^{\frac{d}{d+1}}\sigma_k^{\frac{d}{d+1}}}{\Sigma_K + 2A\frac{1}{\bar{S}^{\frac{d+2}{2(d+1)}}}}(n-K\bar{S}) \\
&\geq \lambda_{K,k}(n-K\bar{S})\Big(1-\frac{2A}{\Sigma_K\bar{S}^{\frac{d+2}{2(d+1)}}}\Big) \\
&\geq \lambda_{K,k}\Big(n-K\bar{S}-\frac{2An}{\Sigma_K\bar{S}^{\frac{d+2}{2(d+1)}}}\Big).
\end{aligned}$$

Using the fact that $\big(\frac{n}{K}\big)^{\frac{d}{d+1}} \geq \bar{S} \geq \Big(\big(\frac{n}{K}\big)^{\frac{1}{d+1}}-1\Big)^d \geq \big(\frac{n}{K}\big)^{\frac{d}{d+1}} - d\big(\frac{n}{K}\big)^{\frac{d-1}{d+1}}$ in the last Equation,

$$\begin{aligned}
C_k &\geq \lambda_{K,k}\Big(n-K\big(\frac{n}{K}\big)^{\frac{d}{d+1}} - \frac{2An}{\Sigma_K}\big(\frac{K}{n}\big)^{\frac{d}{d+1}\times\frac{d+2}{2(d+1)}}\big(1+d\big(\frac{K}{n}\big)^{\frac{1}{d+1}}\big)^{\frac{d+2}{2(d+1)}}\Big) \\
&\geq \lambda_{K,k}\Big(n-K^{\frac{1}{d+1}}n^{\frac{d}{d+1}} - \frac{2An^{\frac{1}{2}+\frac{1}{(d+1)^2}}}{\Sigma_K}K^{\frac{d(d+2)}{2(d+1)^2}}\big(1+[d\big(\frac{K}{n}\big)^{\frac{1}{d+1}}]^{\frac{d+2}{2(d+1)}}\big)\Big) \\
&\geq \lambda_{K,k}\Big(n-(1+2\frac{A}{\Sigma_K}+d\big(\frac{K}{n}\big)^{\frac{d+2}{2(d+1)^2}})K^{\frac{1}{d+1}}n^{\frac{d}{d+1}}\Big), \qquad (10)
\end{aligned}$$

where the last line comes from the fact that $n \geq K$.

We also have

$$C_k - \lfloor C_k^{1/d}\rfloor^d \leq C_k - (C_k^{1/d}-1)^d = C_k\big(1-(1-\frac{1}{C_k^{1/d}})^d\big) \leq dC_k^{\frac{d-1}{d}}.$$

From the last Equation, the definition of S_k and Equation 10 we deduce that (rounding issues)

$$\begin{aligned}
S_k &\geq \max\Big[\bar{S}, C_k\big(1-\frac{d}{C_k^{1/d}}\big)\Big] \\
&\geq \max\Big[\bar{S}, C_k\big(1-\frac{d}{(\bar{S})^{1/d}}\big)\Big] \\
&\geq \max\Big[\bar{S}, \lambda_{K,k}\Big(n-(1+2\frac{A}{\Sigma_K}+d\big(\frac{K}{n}\big)^{\frac{d+2}{2(d+1)^2}})K^{\frac{1}{d+1}}n^{\frac{d}{d+1}}\Big)\Big(1-d\big(\frac{K}{n}\big)^{\frac{1}{d+1}}\Big)\Big] \\
&\geq \max\Big[\bar{S}, \lambda_{K,k}\Big(n-(2+2\frac{A}{\Sigma_K}+d)K^{\frac{1}{d+1}}n^{\frac{d}{d+1}}\Big)\Big].
\end{aligned}$$

We call $N = n-(2+2\frac{A}{\Sigma_K}+d)K^{\frac{1}{d+1}}n^{\frac{d}{d+1}}$ in the sequel. Note that $\forall k$, we have $S_k \geq \max[\bar{S}, \lambda_{K,k}N]$.

Note also that for $\delta \leq 1$, we have

$$\begin{aligned}
A &= (2L\sqrt{d}\sqrt{\log(2K/\delta)})^{\frac{d}{d+1}} \\
&\leq 4(L+1)\sqrt{d}\sqrt{\log(K/\delta)}.
\end{aligned}$$

We thus have that

$$n \geq N \geq n - 7(L+1)d^{3/2}\sqrt{\log(K/\delta)}(1+\frac{1}{\Sigma_K})K^{\frac{1}{d+1}}n^{\frac{d}{d+1}}. \qquad (11)$$

D Proof of Theorem 1

Step 1: Notations Let $\left((\Omega_k^n)_{k\leq K_n}\right)_n$ be a sequence of partitions in hyper-cubic strata of same measure. Let us also assume that the number of strata K_n in partition $(\Omega_k^n)_k$ is such that $\lim_{n\to+\infty} K_n = +\infty$ and $\lim_{n\to\infty} \frac{K_n^{d+2}\log(n)^{d+3}}{n^{d+1}} = 0$. On each of those partitions, $MC-UCB$ is launched with respectively n samples and parameter $\delta_n = \frac{1}{n^2}$.

The number of hyper-cubic sub-strata built by the algorithm in stratum Ω_k^n is $S_{n,k}$. Let us write $\left(\left((\Omega_{k,s}^n)_{s\leq S_{n,k}}\right)_{k\leq K_n}\right)_n$ the partition in hyper-cubic strata formed with those sub-strata. By construction of the algorithm, there is at least one point per sub-stratum. The estimate of the mean of the function is built with the first point in each of those sub-strata.

Let us write $g_n^{(1)}(x) = \sum_{k=1}^{K_n}\sum_{s=1}^{S_{n,k}} \frac{\sigma_{n,k,s}^2}{w_{n,k,s}^{1/2d}}\mathbb{I}\left\{\Omega_{k,s}^n\right\}(x) = \sum_{k=1}^{K_n}\sum_{s=1}^{S_{n,k}} \sigma_{n,k,s}^2 \frac{S_{n,k}^{1/2d}}{w_{n,k}^{1/2d}}\mathbb{I}\left\{\Omega_{k,s}^n\right\}(x)$. From step 1 of the proof of Lemma 1, it converges with n (because $K_n \to +\infty$ when $n\to\infty$ and thus the diameter of each stratum goes to 0) point-wise to $\frac{||\nabla f(x)||_2^2}{12}$.

Let us write $g_n^{(2)}(x) = \sum_{k=1}^{K_n} \frac{\sigma_{n,k}^2}{w_{n,k}^{1/2d}}\mathbb{I}\left\{\Omega_k^n\right\}(x)$. From step 1 of the proof of Lemma 1, it converges with n point-wise to $\frac{||\nabla f(x)||_2^2}{12}$. This convergence implies, as $||\nabla f||_2^2$ is bounded and thus as $\int ||\nabla f||_2^{\frac{d}{d+1}}$ is bounded, by the Theorem of Dominated convergence that $\lim_{n\to+\infty}\Sigma_{K_n} = \lim_{n\to+\infty}\int_{[0,1]^d}(g_n^{(2)}(x))^{\frac{d}{2(d+1)}}dx = \int_{[0,1]^d}(\frac{||\nabla f(x)||_2}{12})^{\frac{d}{(d+1)}}dx > 0$.

Define $\lambda_n(x) = \sum_{k=1}^{K_n}\frac{\lambda_{K_n,k}}{w_{n,k}}\mathbb{I}\left\{\Omega_k^n\right\} = \sum_{k=1}^{K_n}\frac{(w_{n,k}\sigma_{n,k})^{\frac{d}{d+1}}}{w_{n,k}\Sigma_{K_n}}\mathbb{I}\left\{\Omega_k^n\right\} = \frac{(g_n(x))^{\frac{d}{2(d+1)}}}{\Sigma_{K_n}}$. We thus know, as the limit of $(\Sigma_{K_n})_n$ exists and is bigger than 0, that $\lambda_n(x)$ converges pointwise to $s(x) = \frac{||\nabla f(x)||_2^{\frac{d}{d+1}}}{\int_{[0,1]^d}||\nabla f(x)||_2^{\frac{d}{(d+1)}}dx}$.

Let us also define $s_n(x) = \sum_{k=1}^{K_n}\frac{S_{n,k}}{nw_{n,k}}\mathbb{I}\left\{\Omega_k^n\right\}(x)$.

-

Step 1: Majoration of of $\frac{1}{s_n}$. Let us consider only functions f that are not everywhere constant on the domain, as otherwise the bound on the pseudo-risk is trivial[6]. Then $\exists\mathcal{X}\in[0,1]^d$ such that $\mathcal{X}$ is measurable and such that $\int_{\mathcal{X}}1 > 0$, and such that $\forall x\in\mathcal{X}, ||\nabla f(x)||_2 > 0$. Then $\int_{[0,1]^d}(\frac{||\nabla f(x)||_2}{12})^{\frac{d}{(d+1)}}dx > 0$.
Let N_n be defined as in the proof of Lemma 3, i.e. N_n as in Equation 11. As $\lim_{n\to+\infty}\Sigma_{K_n} = \int_{[0,1]^d}(\frac{||\nabla f(x)||_2}{12})^{\frac{d}{(d+1)}}dx$, we know that for any n sufficiently large, $\lim_n\Sigma_{K_n} \geq \frac{1}{2}\int_{[0,1]^d}(\frac{||\nabla f(x)||_2}{12})^{\frac{d}{(d+1)}}dx$. We thus have

$$\begin{aligned} n \geq N_n &\geq n - 7(L+1)d^{3/2}\sqrt{\log(K_n/\delta_n)}(1+\frac{1}{\Sigma_{K_n}})K^{\frac{1}{d+1}}n^{\frac{d}{d+1}} \\ &\geq n - C\sqrt{\log(K_n n^2)}K_n^{\frac{1}{d+1}}n^{\frac{d}{d+1}}, \end{aligned}$$

with $C < +\infty$ as $\int_{[0,1]^d}(\frac{||\nabla f(x)||_2}{12})^{\frac{d}{(d+1)}}dx > 0$. As by definition of the sequence of partitions, $\lim_{n\to+\infty}\sqrt{\log(K_n n^2)}\left(\frac{K_n}{n}\right)^{\frac{1}{d+1}} = 0$, we know that $\lim_{n\to+\infty}\frac{N_n}{n} = 1$.

By Lemma 3, with probability $1-\delta_n$, $\forall k, S_{n,k} \geq \lambda_{K_n,k}N_n$. We thus have

$$\mathbb{P}\left(\frac{1}{s_n(x)} - \frac{1}{\lambda_n(x)} \geq \frac{1}{\lambda_n(x)}(\frac{n}{N_n}-1)\right) \leq \delta_n,$$

[6] If the function is everywhere constant, the samples are always equal to the integral, and the pseudo-risk of the estimate is zero.

which leads to

$$\mathbb{P}\left(\frac{1}{s_n(x)} \geq \frac{1}{\lambda_n(x)}\frac{n}{N_n}\right) \leq \delta_n.$$

Let $\mathcal{X}^+ = \{x \in [0,1]^d : ||\nabla f||_2 > 0\}$. By the last Equation, $\forall \epsilon > 0$, $\forall x \in \mathcal{X}^+$, for n sufficiently large ($\exists n'$ such that $\forall n \geq n'$), $\mathbb{P}(\frac{1}{s_n(x)} - \frac{1}{\lambda_n(x)} \geq \epsilon) \leq \delta_n$. Note that $\sum_{n=1}^{+\infty} \delta_n = \sum_{n=1}^{+\infty} \frac{1}{n^2} \leq +\infty$. We can thus use Borel-Cantelli's Theorem and this gives us that on $\mathcal{X}^+$, $\limsup_n \frac{1}{s_n(x)} - \frac{1}{\lambda_n(x)} \leq 0$ a.s..

We thus deduce (i) by the definition of λ_n and the fact that it converges almost surely to s and (ii) by the fact that $\lim_n \frac{N_n}{n} = 1$, that $\limsup_n \frac{1}{\lambda_n(x)} \leq \frac{1}{s(x)}$ a.s. (since, by definition, $s_n(x) \geq \frac{\bar{S}}{nw_{n,K}} > 0$).

From that we deduce that $\forall x \in \mathcal{X}^+$, $\limsup_n \frac{1}{s_n(x)} \leq \frac{1}{s(x)}$ a.s.. As on $[0,1]^d - \mathcal{X}^+$, $s(x) = 0$, we have $\forall x \in [0,1]^d$, that $\limsup_n \frac{1}{s_n(x)} \leq \frac{1}{s(x)}$ a.s..

Step 2: Convergence rate of the pseudo-risk. The pseudo-risk of the estimate $\hat{\mu}_n$ is

$$\sum_{k=1}^{K_n}\sum_{s=1}^{S_{n,k}}\left(\frac{w_{n,k}}{S_{n,k}}\right)^2 \sigma^2_{n,k,s} = n^{1+2/d}\int_{[0,1]^d} g_n^{(1)}(x)\frac{1}{s_n(x)^{1+2/d}}dx.$$

On $[0,1]^d$, $g_n^{(1)}$ converges pointwise to $\frac{||\nabla f||_2^2}{12}$, and $\limsup_{n\to+\infty} \frac{1}{s_n(x)^{1+2/d}} \leq \frac{1}{s(x)^{1+2/d}}$ a.s. We finally have by Fatou's Lemma that

$$\begin{aligned}
\int_{[0,1]^d} g_n^{(1)}(x)\frac{1}{s_n(x)^{1+2/d}}dx &\leq \int_{[0,1]^d} \limsup_n \left(g_n^{(1)}(x)\frac{1}{s_n(x)^{1+2/d}}\right)dx \\
&\leq \int_{[0,1]^d} \limsup_n g_n^{(1)}(x) \limsup_n \frac{1}{s_n(x)^{1+2/d}}dx \\
&\leq \int_{[0,1]^d} \frac{||\nabla f||_2^2}{12}\frac{1}{s(x)^{1+2/d}}dx.
\end{aligned}$$

By plugging in the last Equation the Definition of s, we conclude the proof.

E Proof of Theorems 2

Step 0: Some inequalities when the second derivative of f is bounded Let a be a point in Ω.

f admits a Taylor expansion in any point. For any $x \in \Omega$ have $|f(x) - f(a) + \nabla f(a).(x-a)| \leq M||x-a||_2^2$ with $2M$ a bound of the second derivative of f.
Note also that $||\nabla f(x) - \nabla f(a)||_2 \leq M||x-a||_2$.

Note also that

$$\begin{aligned}
\left|||\nabla f(x)||_2^2 - ||\nabla f(a)||_2^2\right| &\leq \left|\left(||\nabla f(x)||_2\right)^2 - ||\nabla f(a)||_2^2\right| \\
&\leq \left|\left(||\nabla f(a)||_2 + M||x-a||_2\right)^2 - ||\nabla f(a)||_2^2\right| \\
&\leq \left|||\nabla f(a)||_2^2 + 2M||\nabla f(a)||_2||x-a||_2 + M^2||x-a||_2^2 - ||\nabla f(a)||_2^2\right| \\
&\leq 2M||\nabla f(a)||_2||x-a||_2 + M^2||x-a||_2^2.
\end{aligned}$$

This means that

$$\left|||\nabla f(x)||_2 - ||\nabla f(a)||_2\right| \leq M||x-a||_2. \tag{12}$$

Step 1: Variance on a small box Let us place us on one small box of size w and such that the corresponding domain is $\Omega_w = \prod[a_i - \frac{w^{1/d}}{2}, a_i + \frac{w^{1/d}}{2}]$. We can do a Taylor expansion in a and have

$$|f(x) - f(a) + \nabla f(a)(x-a)| \leq M||x-a||_2^2,$$

with $2M$ a bound of the second derivative of f.

Note that because of the previous equation

$$\begin{aligned}|\frac{1}{w}\int_{\Omega_w}\Big(f(u) - f(a) + \nabla f(a)(u-a)\Big)du| &\leq \frac{1}{w}\int_{\Omega_w}|f(u) - f(a) + \nabla f(a)(u-a)|du \\ &\leq M||x-a||_2^2. \end{aligned} \tag{13}$$

This implies because $a_i = \int_{a_i - \frac{w^{1/d}}{2}}^{a_i + \frac{w^{1/d}}{2}} udu$ that

$$|\frac{1}{w}\int_{\Omega_w} f(u)du - f(a)| \leq M||x-a||_2^2. \tag{14}$$

Finally, by combining Equations 13 and 14, we get

$$|f(x) - \frac{1}{w}\int_{\Omega_w} f(u)du + \nabla f(a)(x-a)| \leq 2M||x-a||_2^2.$$

Triangle inequality on the last Equation leads to

$$|f(x) - \frac{1}{w}\int_{\Omega_w} f(u)du| \leq |\nabla f(a)(x-a)| + 2M||x-a||_2^2.$$

This means by integrating that

$$\begin{aligned}\int_{\Omega_w}\Big(f(x) - \frac{1}{w}\int_{\Omega_w} f(u)du\Big)^2 dx &\leq \int_{\Omega_w}\Big(|\nabla f(a)(x-a)| + 2M||x-a||_2^2\Big)^2 dx \\ &\leq \int_{\Omega_w}\Big(\nabla f(a)(x-a)\Big)^2 dx \qquad (15)\\ &\quad + 2M\int_{\Omega_w}\Big(\nabla f(a)(x-a)|\Big)||x-a||_2^2 dx \qquad (16)\\ &\quad + 4M^2\int_{\Omega_w}||x-a||_2^4 dx. \qquad (17)\end{aligned}$$

Note first that because $a_i = \int_{a_i - \frac{w^{1/d}}{2}}^{a_i + \frac{w^{1/d}}{2}} udu$, we have for the term in Equation 15

$$\begin{aligned}\int_{\Omega_w}\Big(\nabla f(a)(x-a)\Big)^2 dx &= \int_{\Omega_w}\Big(\sum_{i=1}^d \nabla f(a)_i(x_i - a_i)\Big)^2 dx \\ &= w^{1-1/d}\sum_{i=1}^d \int_{a_i - \frac{w^{1/d}}{2}}^{a_i + \frac{w^{1/d}}{2}} \nabla f(a)_i^2 (x_i - a_i)^2 dx_i \\ &= \sum_{i=1}^d \nabla f(a)_i^2 \frac{w^{1+2/d}}{12} \\ &= \frac{w^{1+2/d}}{12}||\nabla f(a)||_2^2. \end{aligned} \tag{18}$$

Now note that for the term in Equation 17

$$\begin{aligned}\int_{\Omega_w}||x-a||_2^4 dx &= \int_{\Omega_w}\Big(\sum_{i=1}^d (x_i - a_i)^2\Big)^2 dx \\ &\leq d^2 w^{1+4/d}. \end{aligned} \tag{19}$$

Now note that because of Cauchy-Schwartz and by using Equations 18 and 19, we have for the term in Equation 16

$$\begin{aligned}\int_{\Omega_w}\Big(\nabla f(a)(x-a)|\Big)||x-a||_2^2dx &\leq \sqrt{\int_{\Omega_w}\Big(\nabla f(a)(x-a)|\Big)^2dx}\sqrt{\int_{\Omega_w}||x-a||_2^4dx}\\ &\leq ||\nabla f(a)||_2 w^{1/2+1/d}\sqrt{d^2w^{1+4/d}}\\ &\leq d||\nabla f(a)||_2 w^{1+3/d}. \end{aligned} \tag{20}$$

We thus have by combining Equations 15, 16, 17, 18, 20 and 19

$$\int_{\Omega_w}\Big(f(x)-\frac{1}{w}\int_{\Omega_w}f(u)du\Big)^2dx \leq \frac{||\nabla f(a)||_2^2}{12}w^{1+2/d}+2Md||\nabla f(a)||_2w^{1+3/d}+4M^2d^2w^{1+4/d}.$$

This leads to using Step 0 in Proof B

$$\begin{aligned} w^2\sigma^2 &\leq \frac{||\nabla f(a)||_2^2}{12}w^{2+2/d}+2Md||\nabla f(a)||_2w^{2+3/d}+4M^2d^2w^{2+4/d}\\ &= w^{2+2/d}\big(\frac{||\nabla f(a)||_2}{2\sqrt{3}}+2Mdw^{1/d}\big)^2. \end{aligned} \tag{21}$$

In the same way, one can prove

$$w^2\sigma^2 \geq w^{2+2/d}\big(\frac{||\nabla f(a)||_2}{2\sqrt{3}}-2Mdw^{1/d}\big)^2. \tag{22}$$

Step 2: Majoration on the strata Lemma 3 tells us that with probability $1-\delta$ (i.e. on the event ξ), each stratum Ω_k is partitioned in $S_k \geq \max\left[\lambda_{p,K}N, \bar{S}\right]$ hyper-cubic substrata $\Omega_{k,i}$ of same measure, and that that there is at least one sample per stratum.The measure of those sub-strata is thus $w_{k,i}=\frac{w_k}{S_k}$.

We have for stratum $\Omega_{k,i}$ by using Equation 21

$$w_{k,i}^2\sigma_{k,i}^2 \leq w_{k,i}^{2+2/d}\big(\frac{||\nabla f(a_{k,i})||_2}{2\sqrt{3}}+2Mdw_{k,i}^{1/d}\big)^2,$$

where $a_{k,i}$ is the center of stratum $\Omega_{k,i}$.

Let $c_{k,i}$ be a point in $\Omega_{k,i}$ such that $c_{k,i}=\arg\min_{c\in\Omega_{k,i}}||\nabla f(c)||_2$. By using that and Equation 12, we get that the variance on strata k that is bounded by

$$\begin{aligned}\sum_{i=1}^{S_k}w_{k,i}^2\sigma_{k,i}^2 &\leq \sum_{i=1}^{S_k}w_{k,i}^{2+2/d}\big(\frac{||\nabla f(a_{k,i})||_2}{2\sqrt{3}}+2Mdw_{k,i}^{1/d}\big)^2\\ &\leq \sum_{i=1}^{S_k}w_{k,i}^{2+2/d}\big(\frac{||\nabla f(c_{k,i})||_2}{2\sqrt{3}}+3Mdw_{k,i}^{1/d}\big)^2\\ &\leq \frac{w_k}{S_k}\sum_{i=1}^{S_k}w_{k,i}^{\frac{d+2}{d}}\big(\frac{||\nabla f(c_{k,i})||_2}{2\sqrt{3}}+3Mdw_{k,i}^{1/d}\big)^2. \end{aligned}$$

Let us call $g(x)=\frac{||\nabla f(x)||_2}{2\sqrt{3}}+3Mdw_k^{1/d}$. As $w_k \geq w_{k,i}$, and $||\nabla f||_2$ is positive, we have

$$\sum_{i=1}^{S_k}w_{k,i}^2\sigma_{k,i}^2 \leq \frac{w_k}{S_k}\sum_{i=1}^{S_k}w_{k,i}^{\frac{d+2}{d}}g(c_{k,i})^2. \tag{23}$$

Step 3: Minoration of the number of sub-strata in each stratum By setting Equation 21 to the power $\frac{d}{2(d+1)}$, we get on stratum Ω_k that

$$(w_k\sigma_k)^{\frac{d}{d+1}} \leq w_k\big(\frac{||\nabla f(a_k)||_2}{2\sqrt{3}} + 2Mdw_k^{1/d}\big)^{\frac{d}{d+1}}.$$

Let c_k^m be a point in Ω_k such that $c_k^m = \arg\min_{c\in\Omega_k} ||\nabla f(c)||_2$. Note that this implies that $\sum_{k=1}^K w_k\big(\frac{||\nabla f(c_k^m)||_2}{2\sqrt{3}} + 3Mdw_k^{1/d}\big)^{\frac{d}{d+1}} \leq \int_{[0,1]^d}\big(\frac{||\nabla f(u)||_2}{2\sqrt{3}} + 3Mdw_k^{1/d}\big)^{\frac{d}{d+1}}du$. By using that and Equation 12, we get that $\Sigma_K = \sum_k (w_k\sigma_k)^{\frac{d}{d+1}}$ is bounded as

$$\begin{aligned}
\Sigma_K &\leq \sum_{k=1}^K w_k\big(\frac{||\nabla f(a_k)||_2}{2\sqrt{3}} + 2Mdw_k^{1/d}\big)^{\frac{d}{d+1}} \\
&\leq \sum_{k=1}^K w_k\big(\frac{||\nabla f(c_k^m)||_2}{2\sqrt{3}} + 3Mdw_k^{1/d}\big)^{\frac{d}{d+1}} \\
&\leq \int_{[0,1]^d}\big(\frac{||\nabla f(u)||_2}{2\sqrt{3}} + 3Mdw_k^{1/d}\big)^{\frac{d}{d+1}}du \\
&\leq \int_{[0,1]^d} g(u)^{\frac{d}{d+1}}du. \qquad (24)
\end{aligned}$$

In the same way, we can deduce

$$\Sigma_K \geq \int_{[0,1]^d}\big(\frac{||\nabla f(u)||_2}{2\sqrt{3}} - 3Mdw_k^{1/d}\big)^{\frac{d}{d+1}}du. \qquad (25)$$

Let c_k^M be a point in Ω_k such that $c_k^M = \arg\max_{c\in\Omega_k} ||\nabla f(c)||_2$. For a stratum k, by using Equations 22 and 12

$$\begin{aligned}
(w_k\sigma_k)^{\frac{d+2}{d+1}} &\geq w_k^{\frac{d+2}{d}}\big(\frac{||\nabla f(a_k)||_2}{2\sqrt{3}} - 2Mdw_k^{1/d}\big)^{\frac{d+2}{d+1}} \\
&\geq w_k^{\frac{d+2}{d}}\big(\frac{||\nabla f(c_k^M)||_2}{2\sqrt{3}} - 3Mdw_k^{1/d}\big)^{\frac{d+2}{d+1}}.
\end{aligned}$$

As for any $u > 0$ and $\alpha > 0$ one has $(1-u)^{-\alpha} \geq 1 + \alpha u$, the last Equation leads to

$$\begin{aligned}
\frac{1}{(w_k\sigma_k)^{\frac{d+2}{d+1}}} &\leq \frac{1}{w_k^{\frac{d+2}{d}}\big(\frac{||\nabla f(c_k^M)||_2}{2\sqrt{3}} + 3Mdw_k^{1/d} - 3Md(w_k^{1/d} + w_k^{1/d})\big)^{\frac{d+2}{d+1}}} \\
&\leq \frac{1}{w_k^{\frac{d+2}{d}}\big(g(c_k^M) - 6Mdw_k^{1/d}\big)^{\frac{d+2}{d+1}}} \\
&\leq \frac{1}{w_k^{\frac{d+2}{d}} g(c_k^M)^{\frac{d}{d+1}}\big(1 - \frac{6Mdw_k^{1/d}}{g(c_k^M)}\big)^{\frac{d+2}{d+1}}} \\
&\leq \frac{1}{w_k^{\frac{d+2}{d}}\big(g(c_k^M)\big)^{\frac{d+2}{d+1}}}\big(1 + (\frac{d+2}{d+1})\frac{6Mdw_k^{1/d}}{g(c_k^M)}\big) \\
&\leq \frac{1}{w_k^{\frac{d+2}{d}}}\big(\frac{1}{\big(g(c_k^M)\big)^{\frac{d+2}{d+1}}} + \frac{9Mdw_k^{1/d}}{(g(c_k^M))^{\frac{2d+3}{d+1}}}\big).
\end{aligned}$$

As $w_{k,i} = \frac{w_k}{S_k}$ this leads with the last Equation and Equation 24

$$(w_{k,i})^{\frac{d+2}{d}} \leq \Big(\frac{\int_{[0,1]^d}\big(g(u)\big)^{\frac{d}{d+1}}du}{N}\Big)^{\frac{d+2}{d}}\big(\frac{1}{\big(g(c_k^M)\big)^{\frac{d+2}{d+1}}} + \frac{9Mdw_k^{1/d}}{(g(c_k^M))^{\frac{2d+3}{d+1}}}\big). \qquad (26)$$

Step 4: Bound on the pseudo-risk As $c_k^M = \max_{c\in\Omega_k} ||\nabla f(c)||_2$ and $c_{k,i} = \min_{c\in\Omega_{k,i}} ||\nabla f(c)||_2$, and as $g(x) = \frac{||\nabla f(x)||_2}{2\sqrt{3}} + 3Mdw_k^{1/d}$, we have for any $(a,b) \geq 0$ that $\frac{g(c_{k,i})^a}{g(c_k^M)^b} \leq \min_{c\in\Omega_{k,i}} g(c)^{a-b}$. By using that and Equations 23 and 26

$$\begin{aligned}
\sum_{i=1}^{S_k} w_{k,i}^2\sigma_{k,i}^2 \leq & \frac{w_k}{S_k}\Big(\frac{\int_{[0,1]^d}\big(g(u)\big)^{\frac{d}{d+1}}du}{N}\Big)^{\frac{d+2}{d}}\sum_{i=1}^{S_k} w_{k,i}^{\frac{d+2}{d}} g(c_{k,i})^2 \\
\leq & \Big(\frac{\int_{[0,1]^d}\big(g(u)\big)^{\frac{d}{d+1}}du}{N}\Big)^{\frac{d+2}{d}}\frac{w_k}{S_k}\sum_{i=1}^{S_k}\big(\frac{1}{\big(g(c_k^M)\big)^{\frac{d+2}{d+1}}} + \frac{9Mdw_k^{1/d}}{(g(c_k^M))^{\frac{2d+3}{d+1}}}\big)g(c_{k,i})^2 \\
\leq & \Big(\frac{\int_{[0,1]^d}\big(g(u)\big)^{\frac{d}{d+1}}du}{N}\Big)^{\frac{d+2}{d}}\frac{w_k}{S_k}\sum_{i=1}^{S_k}\big(\min_{c\in\Omega_{k,i}} g(c)^{\frac{d}{d+1}} + \min_{c\in\Omega_{k,i}}\frac{9Mdw_k^{1/d}}{(g(c))^{\frac{1}{d+1}}}\big).
\end{aligned}$$

Note also that by definition, $g(x) \geq 3Mdw_k^{1/d}$. From that and the previous Equation, we deduce

$$\begin{aligned}
\sum_{i=1}^{S_k} w_{k,i}^2\sigma_{k,i}^2 \leq & \Big(\frac{\int_{[0,1]^d}\big(g(u)\big)^{\frac{d}{d+1}}du}{N}\Big)^{\frac{d+2}{d}}\frac{w_k}{S_k}\sum_{i=1}^{S_k}\big(\min_{c\in\Omega_{k,i}} g(c)^{\frac{d}{d+1}} + \frac{9Mdw_k^{1/d}}{(3Mdw_k^{1/d})^{\frac{1}{d+1}}}\big) \\
\leq & \Big(\frac{\int_{[0,1]^d}\big(g(u)\big)^{\frac{d}{d+1}}du}{N}\Big)^{\frac{d+2}{d}} w_k\big(\frac{1}{w_k}\int_{\Omega_k} g(u)^{\frac{d}{d+1}}du + 9Mdw_k^{\frac{1}{d+1}}\big).
\end{aligned}$$

Finally, by summing over all strata and because all strata have same measure $w_k = \frac{1}{K}$

$$\begin{aligned}
\sum_{i=1}^{K}\sum_{i=1}^{S_k} w_{k,i}^2\sigma_{k,i}^2 \leq & \Big(\frac{\int_{[0,1]^d}\big(g(u)\big)^{\frac{d}{d+1}}du}{N}\Big)^{\frac{d+2}{d}}\sum_{k=1}^{K}\big(\int_{\Omega_k} g(u)^{\frac{d}{d+1}}du + w_k\times 9Mdw_k^{\frac{1}{d+1}}\big) \\
\leq & \Big(\frac{\int_{[0,1]^d}\big(g(u)\big)^{\frac{d}{d+1}}du}{N}\Big)^{\frac{d+2}{d}}\big(\int_{[0,1]^d} g(u)^{\frac{d}{d+1}}du + 9Md\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\big) \\
\leq & \frac{1}{N^{\frac{d+2}{d}}}\Big(\big(\int_{[0,1]^d} g(u)^{\frac{d}{d+1}}du\big)^{\frac{2(d+1)}{d}} + 9Md\big(\int_{[0,1]^d} g(u)^{\frac{d}{d+1}}du\big)^{\frac{d+2}{d}}\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big).
\end{aligned} \tag{27}$$

Step 5: Bound on $\int_{[0,1]^d} g(u)^{\frac{d}{d+1}}du$ Note that because $\frac{d}{d+1} \leq 1$, we have

$$\begin{aligned}
g(u)^{\frac{d}{d+1}} = & \Big(\frac{||\nabla f(u)||_2}{2\sqrt{3}} + 3Mdw_k^{1/d}\Big)^{\frac{d}{d+1}} \\
\leq & \big(\frac{||\nabla f(u)||_2}{2\sqrt{3}}\big)^{\frac{d}{d+1}} + 3Mdw_k^{\frac{1}{d+1}}
\end{aligned}$$

We thus have

$$\int_{[0,1]^d} g(u)^{\frac{d}{d+1}}du \leq \int_{[0,1]^d}\big(\frac{||\nabla f(u)||_2}{2\sqrt{3}}\big)^{\frac{d}{d+1}}du + 3Mdw_k^{\frac{1}{d+1}}. \tag{28}$$

Note also that for $x \geq 0$, and as $\frac{2(d+1)}{d} \leq 4$, we have

$$(1+x)^{\frac{2(d+1)}{d}} \leq (1+x)^4 \leq 1 + 2^4\max(x, x^2, x^3, x^4).$$

Let us call $\Sigma = \int_{[0,1]^d} \big(\frac{||\nabla f(u)||_2}{2\sqrt{3}}\big)^{\frac{d}{d+1}} du$. Then by applying the previous result to Equation 28, we get

$$\begin{aligned}\Big(\int_{[0,1]^d} g(u)^{\frac{d}{d+1}} du\Big)^{\frac{2(d+1)}{d}} &\leq \Big(\int_{[0,1]^d} \big(\frac{||\nabla f(u)||_2}{2\sqrt{3}}\big)^{\frac{d}{d+1}} du + 3Mdw_k^{\frac{1}{d+1}}\Big)^{\frac{2(d+1)}{d}} \\ &= \Sigma^{\frac{2(d+1)}{d}}\Big(1+\frac{3Md}{\Sigma}w_k^{\frac{1}{d+1}}\Big)^{\frac{2(d+1)}{d}} \\ &\leq \Sigma^{\frac{2(d+1)}{d}} + 16\Sigma^{\frac{2(d+1)}{d}}\Big(1+\frac{3Md}{\Sigma}\Big)^4 w_k^{\frac{1}{d+1}}. \end{aligned} \tag{29}$$

Note also that by Equation 12, we know that $||\nabla f(u)||_2 \leq ||\nabla f(0)||_2 + M\sqrt{d}$. From that we deduce that

$$\begin{aligned}\int_{[0,1]^d} g(u)^{\frac{d}{d+1}} du &\leq \Sigma + 3Mdw_k^{\frac{1}{d+1}} \\ &\leq \Sigma + 3Md. \end{aligned} \tag{30}$$

Step 6: Final bound on the pseudo-risk From Equations 27, 29 and 30, we deduce

$$\begin{aligned}\sum_{i=1}^K\sum_{i=1}^{S_k} w_{k,i}^2\sigma_{k,i}^2 \leq& \frac{1}{N^{\frac{d+2}{d}}}\Big(\big(\int_{[0,1]^d} g(u)^{\frac{d}{d+1}} du\big)^{\frac{2(d+1)}{d}} + 9Md\big(\int_{[0,1]^d} g(u)^{\frac{d}{d+1}} du\big)^{\frac{d+2}{d}}\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big) \\ \leq& \frac{1}{N^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + 16\Sigma^{\frac{2(d+1)}{d}}\Big(1+\frac{3Md}{\Sigma}\Big)^4 w_k^{\frac{1}{d+1}} \\ &+ 9Md\big(\Sigma+3Md\big)^{\frac{d+2}{d}}\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big] \\ \leq& \frac{1}{N^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + 25Md(\Sigma+1)^{\frac{2(d+1)}{d}}\Big(1+\frac{3Md}{\Sigma}\Big)^4\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big] \\ \leq& \frac{1}{N^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + C\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big],\end{aligned}$$

where $C = 25Md(\Sigma+1)^{\frac{2(d+1)}{d}}\Big(1+\frac{3Md}{\Sigma}\Big)^4$.

Note that $N = n - (2+2\frac{A}{\Sigma_K}+d)K^{\frac{1}{d+1}}n^{\frac{d}{d+1}} = n - BK^{\frac{1}{d+1}}n^{\frac{d}{d+1}}$, where $B = 2+2\frac{A}{\Sigma_K}+d$. From plugging that in the last Equation, we get

$$\begin{aligned}\sum_{i=1}^K\sum_{i=1}^{S_k} w_{k,i}^2\sigma_{k,i}^2 \leq& \frac{1}{\Big(n - BK^{\frac{1}{d+1}}n^{\frac{d}{d+1}}\Big)^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + C\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big] \\ \leq& \frac{1}{n^{\frac{d+2}{d}}\Big(1 - BK^{\frac{1}{d+1}}n^{-\frac{1}{d+1}}\Big)^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + C\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big] \\ \leq& \frac{1}{n^{\frac{d+2}{d}}}\Big[1+(\frac{d+2}{d})BK^{\frac{1}{d+1}}n^{-\frac{1}{d+1}}\Big]\Big[\Sigma^{\frac{2(d+1)}{d}} + C\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big] \\ \leq& \frac{1}{n^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + 3\Sigma^{\frac{2(d+1)}{d}}BK^{\frac{1}{d+1}}n^{-\frac{1}{d+1}} + C\big(\frac{1}{K}\big)^{\frac{1}{d+1}} + 3BCn^{-\frac{1}{d+1}}\Big],\end{aligned}$$

where we use for passing from the second to the third line of the Equation that $(1-u)^{-\alpha} \leq 1+\alpha u$.

By it's definition, $C \geq \Sigma^{\frac{2(d+1)}{d}}$ and this leads to

$$\sum_{i=1}^K\sum_{i=1}^{S_k} w_{k,i}^2\sigma_{k,i}^2 \leq \frac{1}{n^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + 6BCK^{\frac{1}{d+1}}n^{-\frac{1}{d+1}} + C\big(\frac{1}{K}\big)^{\frac{1}{d+1}}\Big]. \tag{31}$$

Note first that by Equation 25 and because $||\nabla f||_2 \leq L$ we have

$$\begin{aligned}\Sigma_K \geq & \int_{[0,1]^d} \big(\frac{||\nabla f(u)||_2}{2\sqrt{3}} - 3Mdw_k^{1/d}\big)^{\frac{d}{d+1}} du \\ \geq & \Sigma - 3LMdw_k^{\frac{1}{d+1}}.\end{aligned}$$

From that we deduce that

$$\begin{aligned}B &\leq 2 + 2\frac{4(L+1)\sqrt{d}\sqrt{\log(K/\delta)}}{\Sigma - 3LMdw_k^{\frac{1}{d+1}}} + d \\ &\leq 2 + 8\frac{(L+1)\sqrt{d}\sqrt{\log(K/\delta)}}{\Sigma} + 2LMdw_k^{\frac{1}{d+1}}\frac{(L+1)\sqrt{d}\sqrt{\log(K/\delta)}}{\Sigma^2} + d \\ &\leq 10(L+1)\sqrt{d}\sqrt{\log(K/\delta)}(1 + \frac{1}{\Sigma^2}).\end{aligned}$$

By plugging in Equation 31 the definition of C and the bound on B computed above, we obtain

$$\begin{aligned}\sum_{i=1}^{K}\sum_{i=1}^{S_k} w_{k,i}^2\sigma_{k,i}^2 \leq & \frac{1}{n^{\frac{d+2}{d}}}\Big[\Sigma^{\frac{2(d+1)}{d}} + 650M(L+1)d^{3/2}\Big(1 + \frac{3Md}{\Sigma}\Big)^4\sqrt{\log(K/\delta)}K^{\frac{1}{d+1}}n^{-\frac{1}{d+1}} \\ & + 25Md(\Sigma+1)^{\frac{2(d+1)}{d}}\Big(1 + \frac{3Md}{\Sigma}\Big)^4(\frac{1}{K})^{\frac{1}{d+1}}\Big].\end{aligned}$$

This concludes the proof.