Hardy's Tauberian theorem, bounded variation and Fourier series

MICHAEL MÜGER

May 25, 2010

1 Cesàro summation and Hardy's theorem

1.1 DEFINITION *Given a sequence* $(a_n \in \mathbb{C})_{n\in\mathbb{N}_0}$, *we define*

$$s_n = \sum_{k=0}^{n} a_n, \qquad \sigma_n = \frac{\sum_{k=0}^{n} s_n}{n+1}.$$

If $\lim_{n\to\infty} \sigma_n = A \in \mathbb{C}$ *we say that "*$\sum_k a_k$ *is Cesàro summable to* A*" and write*

$$C\!-\!\sum_{k=0}^{\infty} a_n = A.$$

1.2 REMARK The following facts are easy to show:

(a) $\lim_{n\to\infty} s_N = A$ implies $\lim_{n\to\infty} \sigma_N = A$. (Ordinary convergence implies Cesàro convergence.)

(b) There are sequences (a_k) such that $C\!-\!\sum_{k=0}^{\infty} a_n$ exists but $\sum_{k=0}^{\infty} a_n$ does not.

(c) If $C\!-\!\sum_{k=0}^{\infty} a_n = A$ **and** $a_n = o(\frac{1}{n})$ then $\sum_{k=0}^{\infty} a_n = A$.

The following theorem of Hardy (1910) is better than (c) above since $a_n = O(\frac{1}{n})$ (meaning $|na_n| \le C$ for some $C > 0$ and all $n > 0$) is a weaker assumption than $a_n = o(\frac{1}{n})$ (meaning $\lim_{n\to 0} na_n = 0$).

1.3 Theorem *If* $C-\sum_{k=0}^{\infty} a_n = A$ **and** $a_n = O(\frac{1}{n})$ *then* $\sum_{k=0}^{\infty} a_n = A$.

Proof. Recall that

$$S_n = \sum_{k=0}^{n} a_n, \qquad \sigma_n = \sum_{k=0}^{n} \left(1 - \frac{k}{n+1}\right) a_n.$$

Picking $\lambda > 1$ and recalling that $\lfloor x \rfloor := \max\{n \in \mathbb{Z} \mid n \leq x\}$ we claim that

$$S_n = \frac{\lfloor \lambda n \rfloor + 1}{\lfloor \lambda n \rfloor - n} \left(\sigma_{\lfloor \lambda n \rfloor} - \sum_{n < k \leq \lfloor \lambda n \rfloor} \left(1 - \frac{k}{\lfloor \lambda n \rfloor + 1}\right) a_n \right) - \frac{n+1}{\lfloor \lambda n \rfloor - n} \sigma_n. \tag{1.1}$$

To see this, we observe that

$$\sigma_{\lfloor \lambda n \rfloor} - \sum_{n < k \leq \lfloor \lambda n \rfloor} \left(1 - \frac{k}{\lfloor \lambda n \rfloor + 1}\right) a_n = \sum_{k=0}^{n} \left(1 - \frac{k}{\lfloor \lambda n \rfloor + 1}\right) a_n,$$

thus the first half of the expression equals

$$\frac{\lfloor \lambda n \rfloor + 1}{\lfloor \lambda n \rfloor - n} \sum_{k=0}^{n} \left(1 - \frac{k}{\lfloor \lambda n \rfloor + 1}\right) a_n = \sum_{k=0}^{n} \frac{\lfloor \lambda n \rfloor + 1 - k}{\lfloor \lambda n \rfloor - n} a_n,$$

and subtracting the last term, namely

$$\frac{n+1}{\lfloor \lambda n \rfloor - n} \sigma_n = \frac{n+1}{\lfloor \lambda n \rfloor - n} \sum_{k=0}^{n} \left(1 - \frac{k}{n+1}\right) a_n = \sum_{k=0}^{n} \frac{n+1-k}{\lfloor \lambda n \rfloor - n} a_n,$$

we get

$$\sum_{k=0}^{n} \frac{(\lfloor \lambda n \rfloor + 1 - k) - (n+1-k)}{\lfloor \lambda n \rfloor - n} a_n = \sum_{k=0}^{n} \frac{\lfloor \lambda n \rfloor - n}{\lfloor \lambda n \rfloor - n} a_n = S_n,$$

proving (1.1). If we now let $n \to \infty$ in (1.1) then $\sigma_n \to A$ and $\sigma_{\lfloor \lambda n \rfloor} \to A$ by the assumption of Cesàro summability. Therefore,

$$\frac{\lfloor \lambda n \rfloor + 1}{\lfloor \lambda n \rfloor - n} \sigma_{\lfloor \lambda n \rfloor} - \frac{n+1}{\lfloor \lambda n \rfloor - n} \sigma_n \longrightarrow \frac{\lambda}{\lambda - 1} A - \frac{1}{\lambda - 1} A = A.$$

Thus, (1.1) implies $\lim_{n\to\infty} S_n = A$, provided we can show that the remaining term in (1.1) tends to zero as $n \to \infty$. It is given by

$$\frac{\lfloor \lambda n\rfloor + 1}{\lfloor \lambda n\rfloor - n} \cdot \left| \sum_{n<k\le\lfloor \lambda n\rfloor} \left(1 - \frac{k}{\lfloor \lambda n\rfloor + 1}\right) a_n \right| = \left| \sum_{n<k\le\lfloor \lambda n\rfloor} \frac{\lfloor \lambda n\rfloor + 1 - k}{\lfloor \lambda n\rfloor - n}\, a_n \right|$$

$$\le \sum_{n<k\le\lfloor \lambda n\rfloor} \left| \frac{\lfloor \lambda n\rfloor + 1 - k}{\lfloor \lambda n\rfloor - n} a_n \right| \le \sum_{n<k\le\lfloor \lambda n\rfloor} |a_n|, \tag{1.2}$$

where we used that

$$\left| \frac{\lfloor \lambda n\rfloor + 1 - k}{\lfloor \lambda n\rfloor - n} \right| \le 1$$

for $n < k \le \lfloor \lambda n\rfloor$. Finally using the assumption $a_n = O(\frac{1}{n})$, or $|a_n| \le C/n\ \forall n \ge 1$, we continue the estimate (1.2) as follows:

$$\le \sum_{n<k\le\lfloor \lambda n\rfloor} \frac{C}{n} \le \int_n^{\lfloor \lambda n\rfloor} \frac{C}{n} \le \int_n^{\lambda n} \frac{C}{n} = C(\ln(\lambda n) - \ln n) = C \ln \frac{\lambda n}{n} = C \ln \lambda.$$

(In comparing the sum with the integral, we have used that C/n is monotonously decreasing.) Thus, for any given $\varepsilon > 0$, we can choose $\lambda > 1$ sufficiently close to 1 to make $C \ln \lambda$, and thereby the term with $\sum_{n<k\le\lfloor \lambda n\rfloor}$ smaller than ε, uniformly in n. This proves $S_n \to A$. □

1.4 Remark 1. I thank Lawrence Forooghian from Cambridge University for drawing my attention to a mistake in a previous version of these notes and for providing a correction.

2. Recall the notion of Abel summability: If $(a_n)_{n\in\mathbb{N}_0}$ is such that

$$f(x) = \sum_{k=0}^{\infty} a_k x^k$$

converges for all $|x| < 1$ and $A = \lim_{x\nearrow 1} f(x)$ exists, then A is called the Abel-sum $A\text{–}\sum_{k=0}^{\infty} a_k$. Abel proved that if A=$\sum_{k=0}^{\infty} a_n$ exists then $A\text{–}\sum_{k=0}^{\infty} a_k = A$. As for Cesàro summation, the converse is false. In 1897, Tauber proved

$$A = A\text{–}\sum_{k=0}^{\infty} a_k \ \ \textbf{and}\ \ a_n = o\left(\frac{1}{n}\right) \quad \Rightarrow \quad \sum_{k=0}^{\infty} a_k = A.$$

Since then, a "Tauberian theorem" is a theorem the the effect that summability w.r.t. some summation method together with a decay condition on the coefficients implies summability w.r.t. some weaker method (for example ordinary convergence). Fact (c) of Remark 1.2 and Hardy's Theorem 1.3 are such theorems. Another example: Littlewood proved in 1911 that $o(\frac{1}{n})$ in Tauber's original theorem can be replaced by $O(\frac{1}{n})$. (This is a good deal more difficult to prove than Theorem 1.3, which it implies!)

2 Application to Fourier series

Let $f \in \mathcal{R}[0, 2\pi]$ and define

$$c_n(f) = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx} dx, \qquad S_N(f)(x) = \sum_{n=-N}^{N} c_n(f) e^{inx}.$$

The convergence of $S_N(f)(x)$ to $f(x)$ is a tricky problem, but the Cesàro means

$$\sigma_N(f)(x) = \frac{\sum_{k=0}^{N} S_N(f)(x)}{N+1}$$

of the partial sums $S_N(f)$ behave much better: If f is continuous at x then $\sigma_N(f)(x) \to f(x)$. Furthermore, if f is continuous on $E \subset S^1$ then $\sigma_N(f) \rightrightarrows f$ on E (uniform convergence).

We are now in a position to apply Hardy's theorem to the theory of Fourier series:

2.1 Theorem *Let $f \in \mathcal{R}[0, 2\pi]$ be such that $c_n(f) = O(\frac{1}{|n|})$. Then, as $N \to \infty$ we have*

(a) $S_N(f)(x) \to f(x)$ at every point of continuity of f.

(b) If $f \in C(S^1)$ then $S_N(f) \rightrightarrows f$ (uniform convergence).

Proof. (a) Assume first that f is continuous at 0. We have

$$S_N(f)(0) = \sum_{n=-N}^{N} c_n = c_0 + \sum_{n=1}^{N} (c_n + c_{-n}) = \sum_{n=0}^{N} a_n,$$

where $a_0 = c_0,\ a_n = c_n + c_{-n}$ for $n \geq 1$. Now, by Fejér's theorem, $C\text{–}\sum_{n=0}^{\infty} a_n$ exists (and is equal to $f(0)$). Since $c_n = O(\frac{1}{|n|})$ clearly implies $a_n = O(\frac{1}{n})$, Theorem 1.3 gives that

$$\lim_{N\to\infty} S_N(f)(0) = \lim_{N\to\infty} \sum_{n=-N}^{N} c_n = \sum_{n=0}^{\infty} a_n = f(0).$$

Considering now $f_{x_0}(x) = f(x + x_0)$, we have $c_n(f_{x_0}) = e^{-inx_0} c_n(f)$ and thus $c_n(f_{x_0}) = O(\frac{1}{|n|})$. Thus, if f is continuous at x_0 then the above implies

$$S_N(f)(x_0) = S_N(f_{x_0})(0) \longrightarrow f_{x_0}(0) = f(x_0).$$

(b) A continuous function on S^1 is the same as a continuous periodic function on $\mathbb{R}$. Such a function is uniformly continuous, i.e. for every $\varepsilon > 0$ there is $\delta > 0$ such that $|x - y| < \delta \ \Rightarrow |f(x) - f(y)| < \varepsilon$. Using this, the convergence in (a) is easily seen to be uniform in x. □

2.2 Remark Fejérs theorem generalizes to the situation where f is not continuous at x, but the limits $f(x+)$ and $f(x-)$ both exist, giving $\sigma_N(f)(x) \to \frac{f(x+)+f(x-)}{2}$. Combining this with Hardy's theorem, we see that also (a) of Theorem 2.1 generalizes accordingly.

We are now left with the problem of identifying a natural class of functions for which $c_n(f) = O(\frac{1}{|n|})$.

3 Functions of bounded variation

3.1 Definition *The total variation* $\mathrm{Var}_{[a,b]}(f) \in [0, \infty]$ *of a function* $f : [a,b] \to \mathbb{C}$ *is defined by*

$$\mathrm{Var}_{[a,b]}(f) = \sup_P \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|,$$

where the suprenum is over the partitions $P = \{a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\}$ *of* $[a,b]$. *If* $\mathrm{Var}_{[a,b]}(f) < \infty$ *the* f *has bounded variation on* $[a,b]$.

3.2 Proposition *If* $f : [0, 2\pi] \to \mathbb{C}$ *has bounded variation then*

$$|c_n(f)| \leq \frac{\pi}{2} \frac{\mathrm{Var}_{[0,2\pi]}(f)}{|n|} \quad \forall n \in \mathbb{Z}\backslash\{0\}.$$

Proof. Extending f to a 2π-periodic function on $\mathbb{R}$, we have for $n \in \mathbb{Z}$:

$$c_n(T_a f) = \frac{1}{2\pi}\int_0^{2\pi} f(x+a)e^{-inx}dx = \frac{e^{ina}}{2\pi}\int_0^{2\pi} f(x)e^{-inx}dx = e^{ina}c_n(f).$$

For $n \neq 0$ and $a = \pi/n$ this gives $c_n(T_{\pi/n}f) = -c_n(f)$ and thus

$$c_n(f) = \frac{1}{2}(c_n(f) - c_n(T_{\pi/n}f)) = \frac{1}{2}c_n(f - T_{\pi/n}f),$$

implying

$$|c_n(f)| \leq \frac{1}{2}|c_n(f - T_{\pi/n}f)| \leq \frac{1}{2}\|f - T_{\pi/n}f\|_1. \tag{3.1}$$

(Note that this inequality holds for all f with $\|f\|_1 < \infty$.) Now,

$$\begin{aligned}
\|f - T_{\pi/n}f\|_1 &= \int_0^{2\pi} \left| f(x) - f(x+\frac{\pi}{n})\right| dx \\
&= \sum_{k=1}^{2n} \int_{(k-1)\pi/n}^{k\pi/n} \left| f(x) - f(x+\frac{\pi}{n})\right| dx \\
&= \sum_{k=1}^{2n} \int_0^{\pi/n} \left| f(x+\frac{k-1}{n}\pi) - f(x+\frac{k}{n}\pi)\right| dx \\
&= \int_0^{\pi/n} \sum_{k=1}^{2n} \left| f(x+\frac{k-1}{n}\pi) - f(x+\frac{k}{n}\pi)\right| dx \\
&\leq \int_0^{\pi/n} \mathrm{Var}_{[x,x+2\pi]}(f)\, dx \\
&= \frac{\pi}{n}\mathrm{Var}_{[0,2\pi]}(f).
\end{aligned}$$

(In the last step we have used that f is 2π-periodic.) Together with (3.1) this implies the proposition. □

Combining Proposition 3.2 and Theorem 2.1, we finally have:

3.3 Theorem (Dirichlet-Jordan) *Let $f : [0, 2\pi] \to \mathbb{C}$ have bounded variation. Then*

(a) $\lim_{N\to\infty} S_N(f)(x) = \frac{f(x+) + f(x-)}{2}$ $\forall x \in S^1$. *(Recall that a function of bounded variation automatically has left and right limits $f(x+), f(x-)$ in all points and is Riemann integrable, so that we can define the coefficients $c_n(f)$.*

(b) If $f \in C(S^1)$ has bounded variation then $S_n(f) \rightrightarrows f$.

3.4 Remark Assume that there is a partition P of $[0, 2\pi]$ such that f is continuous and monotonous on each interval $(x_{i-1}, x_i),\ i = 1, \ldots, n$ and the limits $f(x_i+), f(x_i-)$ exist. Then

$$\mathrm{Var}_{[0,2\pi]}(f) = \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| < \infty,$$

and the Theorem applies. This is the case proven by Dirichlet in 1828.

3.5 Remark If $f \in C^1([a, b])$ then

$$\mathrm{Var}_{[a,b]}(f) \leq \int_a^b |f'(x)| dx < \infty.$$

4 Summary of our main results

1. For all $f \in \mathcal{R}[0, 2\pi]$ the formula of Parseval holds:

$$\sum_{n\in\mathbb{Z}} |c_n(f)|^2 = \frac{1}{2\pi} \int_0^{2\pi} |f(x)|^2 dx \ < \infty.$$

   This implies the Riemann-Lebesgue Lemma: $\lim_{|n|\to\infty} c_n(f) = 0$ or $c_n(f) = o(1)$.

2. Defining

$$S_N(f)(x) = \sum_{k=-N}^{N} c_k(f) e^{ikx},$$

   we have

$$\|f - S_N(f)\|_2^2 = \frac{1}{2\pi} \int_0^{2\pi} |f(x) - S_N(f)(x)|^2 dx \to 0.$$

   Note that a priori this implies nothing about pointwise convergence, since there is a sequence $\{f_n\}$ such that $\|f_n\|_2 \to 0$, while $\lim_{n\to\infty} f_n(x)$ exists for **no** x. (However, Fourier series cannot be that badly behaved, at least if f is Riemann integrable. See 15-17 below.)

3. If $f \in C^k(S^1)$ then $c_n(f^{(k)}) = (in)^k c_n(f)$ and $c_n(f^{(k)}) = o(1)$ imply
$$|c_n(f)| = o\left(\frac{1}{|n|^k}\right).$$

4. Thus: If $f \in C^2(S)$ then $\widehat{f}(n) = o(n^{-2})$, thus $\sum_{n\in\mathbb{Z}} |c_n(f)| < \infty$, thus $S_N(f) \rightrightarrows f$ even absolutely!

5. If $f \in C^1(S^1)$ then a combination of Parseval's formula, the Cauchy-Schwarz inequality and $\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$ implies:
$$\|\widehat{f}\|_1 := \sum_{n\in\mathbb{Z}} |\widehat{f}(n)\| \leq \|f\|_1| \frac{\pi}{\sqrt{3}} \|f'\|_2,$$
thus $S_N(f) \to f$ absolutely and uniformly.

6. Fejér: The Fejér sums
$$F_N(f)(x) = \frac{\sum_{k=0}^{N} S_N(f)(x)}{N+1} = \sum_{k=-N}^{N} c_k(f)\left(1 - \frac{|k|}{N+1}\right) e^{ikx}$$
converge to $(f(x+) + f(x-))/2$ whenever $f(x+)$ and $f(x-)$ exist, thus to $f(x)$ at every point x of continuity.

7. If $f \in C(S^1)$ then $F_N(f) \rightrightarrows f$ (uniform convergence). Thus we can uniquely reconstruct f from its Fourier coefficients $\{c_n(f)\}$ even if the Fourier series $S_N(f)$ behaves badly.

8. If $f(x+), f(x-)$ exist and $S_N(f)(x) \to A \in \mathbb{C}$ then $A = (f(x+) + f(x-))/2$. Thus: If the Fourier series converges, it converges to the only reasonable value. (We clearly cannot expect $S_N(f)(x) \to f(x)$ at a discontinuity, since the value of $f(x)$ can be chosen arbitrarily without influencing the coefficients $c_n(f)$.)

9. If $f \in \mathrm{Lip}^{\alpha}[0, 2\pi]$ then $|c_n(f)| = O(|n|^{-\alpha})$.

10. If f has bounded variation then $|c_n(f)| = O(|n|^{-1})$.

11. Dirichlet-Jordan: If f has bounded variation then $S_N(f)(x) \to (f(x+) + f(x-))/2$ everywhere.

12. Special case (Dirichlet): f is piecewise continuous and monotonous.

13. If $f \in C(S^1)$ has bounded variation then $S_N(f) \rightrightarrows f$. In particular this holds for $f \in C^1(S^1)$.

14. Dini: If $f(x+), f(x-)$ exist and, for some $\delta > 0$

$$\int_0^\delta \left| \frac{f(x+t) - f(x+) + f(x-t) - f(x-)}{t} \right| < \infty,$$

then $S_N(f)(x) \to (f(x+)+f(x-))/2$. Note that this is a *local* criterion, whereas the previous assumptions on f were global, i.e. concerned all $x \in S^1$.

15. The preceding condition is satisfied if $|f(x+t) - f(x)| \leq Ct^\alpha$ for some $\alpha > 0$ and t in some neighborhood of x. In particular: when f is differentiable at x.

16. Riemann localization principle: The convergence of $S_N(f)(x)$ depends only on the behavior of f on some neighborhood of x. More precisely: If f, g coincide on some open neighborhood of x then either $S_N(f)(x)$ and $S_N(g)(x)$ both diverge or they converge to the same value.

Here a few important and/or useful facts that we haven't proven:

15. There exist $f \in C(S^1)$ such that $S_N(f)(x)$ is divergent for some x. In fact, for every $E \subset S^1$ of measure zero one can find a function f such that $\lim_{N\to\infty} S_N(f)(x)$ diverges for all $x \in E$. (Notice that a set of measure zero can be dense in S^1.) However, it cannot get worse, as the following result shows.

16. Carleson (1966): If $f \in \mathcal{R}[0, 2\pi]$ then $S_N(f)(x) \to f(x)$ almost everywhere (i.e., on the complement of a set of measure zero). In fact this conclusion holds for any function $f \in \mathcal{L}_p([0, 2\pi])$, i.e. f is "measurable" and
$$\int_0^{2\pi} |f(x)|^p dx < \infty,$$
for some $p > 1$. (Such a function can be unbounded and very discontinuous!)

17. On the other hand, Kolmogorov constructed a function $f \in \mathcal{L}_1([0,2\pi])$, thus f is measuable and

$$\int_0^{2\pi} |f(x)|dx < \infty,$$

such that $\lim_{N\to\infty} S_N(f)(x)$ exists for **no** x.

A Alternative proof of Hardy's theorem

We will use the following *discrete Taylor formula*:

A.1 Lemma *Given a real series* $(a_n)_{n\in\mathbb{N}_0}$*, we define* s_n *as above and* $t_n = \sum_{k=0}^n s_k$*. Then for all* $n, h \in \mathbb{N}_0$ *we have*

$$t_{n+h} = t_n + hs_n + \frac{1}{2}h(h+1)\xi, \tag{A.1}$$

where

$$\min_{n<k\leq n+h} a_k \leq \xi \leq \max_{n<k\leq n+h} a_k. \tag{A.2}$$

Proof. By definition of s_k and t_k we have

$$\begin{aligned} t_{n+h} &= t_n + (s_{n+1} + \cdots + s_{n+h}) \\ &= t_n + hs_n + ha_{n+1} + (h-1)a_{n+2} + \cdots + 2a_{n+h-1} + a_{n+h}. \end{aligned}$$

Now,

$$\begin{aligned} &ha_{n+1} + (h-1)a_{n+2} + \cdots + 2a_{n+h-1} + a_{n+h} \\ &\qquad \leq (h + (h-1) + \cdots + 2 + 1) \max_{n<k\leq n+h} a_k = \frac{n(n+1)}{2} \max_{n<k\leq n+h} a_k, \end{aligned}$$

and similarly

$$\frac{n(n+1)}{2} \min_{n<k\leq n+h} a_k \leq ha_{n+1} + (h-1)a_{n+2} + \cdots + 2a_{n+h-1} + a_{n+h}.$$

Thus

$$t_{n+h} - t_n - hs_n \in \frac{n(n+1)}{2} [\min_{n<k\leq n+h} a_k, \max_{n<k\leq n+h} a_k],$$

and we are done. $\square$

Proof of Theorem 1.3. We may clearly assume that $A = 0$. (Otherwise replace a_0 by $a_0 - A$. This entails that s_n and σ_n are replaced by $s_n - A$ and $\sigma_n - A$.) Furthermore, considering real and imaginary parts separately, it is sufficient to give a proof for real sequences (a_n).

In view of $\sigma_n = t_n/(n+1)$, the assumption $\sigma_n \to 0$ is equivalent to $\frac{t_n}{n} \to 0$. Thus, for every $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that $n \geq N$ implies $|t_n| \leq n\varepsilon$. Solving (A.1) for s_n we have

$$s_n = \frac{t_{n+h} - t_n}{h} - \frac{(h+1)\xi}{2},$$

where, using (A.2) and the assumption $a_n = O(\frac{1}{n})$ in the form $|a_n| \leq C/n$,

$$-\frac{C}{n} \leq \min_{n<k\leq n+h} a_k \leq \xi \leq \max_{n<k\leq n+h} a_k \leq \frac{C}{n}.$$

Thus

$$|s_n| \leq \frac{|t_{n+h}|}{h} + \frac{|t_n|}{h} + \frac{(h+1)C}{2n}.$$

With $|t_n| \leq n\varepsilon$ for $n \geq N$ we have

$$\begin{aligned} |s_n| &\leq \frac{(n+h)\varepsilon}{h} + \frac{n\varepsilon}{h} + \frac{(h+1)C}{2n} = \frac{(2n+h)\varepsilon}{h} + \frac{(h+1)C}{2n} \\ &= \varepsilon + \frac{C}{2n} + \frac{2n\varepsilon}{h} + \frac{hC}{2n}. \end{aligned} \tag{A.3}$$

We now try to minimize this expression by chosing $h \in \mathbb{N}$ cleverly. The minimum of $f(h) = \frac{2n\varepsilon}{h} + \frac{hC}{2n}$ is obtained at the solution of $f'(h) = 0$:

$$-\frac{2n\varepsilon}{h^2} + \frac{C}{2n} = 0 \qquad \Rightarrow \qquad h_{min} = 2n\sqrt{\frac{\varepsilon}{C}}.$$

Now

$$f(h_{min}) = \frac{2n\varepsilon}{h_{min}} + \frac{h_{min}C}{2n} = \frac{2n\varepsilon}{2n\sqrt{\frac{\varepsilon}{C}}} + \frac{2n\sqrt{\frac{\varepsilon}{C}}C}{2n} = 2\sqrt{C\varepsilon}.$$

Of course, h_{min} has no reason to be in $\mathbb{N}$. Defining $h = \lceil h_{min} \rceil$, to wit the smallest natural number $h \geq h_{min}$, we have

$$\frac{2n\varepsilon}{h} + \frac{hC}{2n} \leq 2\sqrt{C\varepsilon} + \frac{C}{2n},$$

since the first term in f can only decrease when we replace h_{min} by h, whereas the second can increase by at most $C/2n$. Plugging this into (A.3), we can conclude

$$\forall \varepsilon > 0 \; \exists N : n \geq N \;\Rightarrow\; |s_n| \leq \varepsilon + \frac{C}{n} + 2\sqrt{C\varepsilon},$$

implying $s_n \to 0$. □