Supplementary Information for *AI without networks*, by P. P. Mitra and C. Sire

Legends of Supplementary Movies

Movie S1: 2-minute collective dynamics of 2 Hilbert agents for a memory $M = 2$, and enforcing the presence of the wall by means of the rejection procedure.

Movie S2: 2-minute collective dynamics of 2 Hilbert agents for a memory $M = 2$, without enforcing the presence of the wall. Note the occurrence of small and short excursions of the fish outside the limits of the tank.

Movie S3: 1-minute collective dynamics of 5 Hilbert agents for a memory $M = 2$, and enforcing the presence of the wall by means of the rejection procedure.

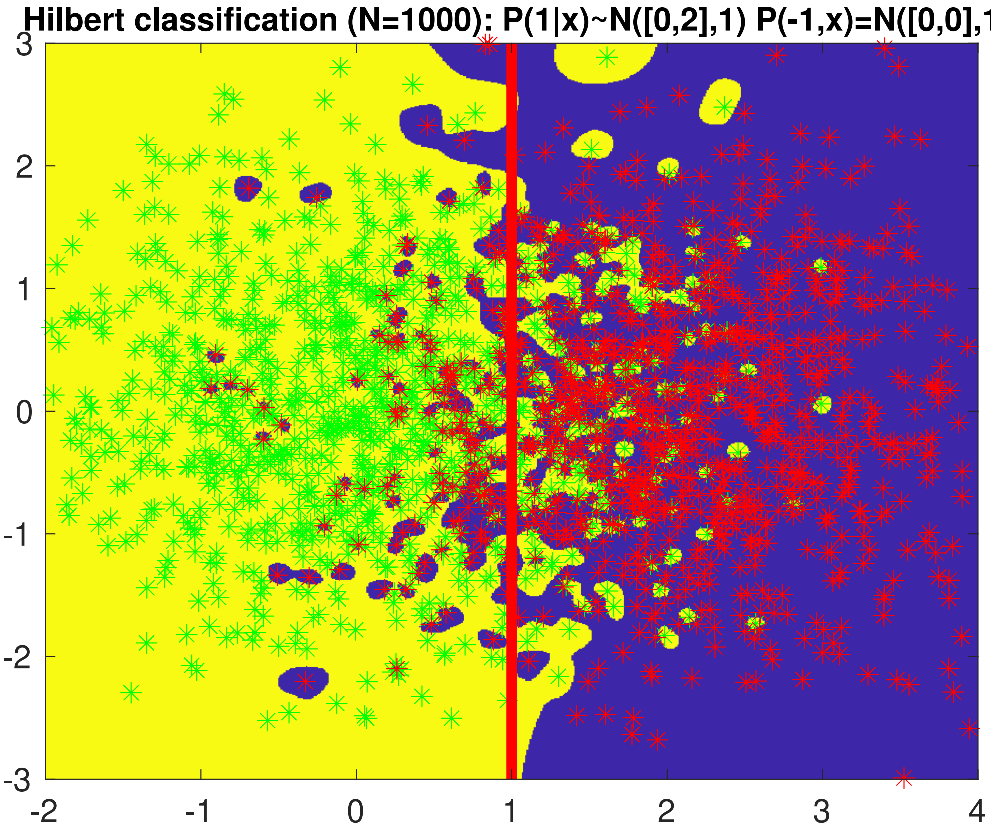
Extended Data Table

$N = 2$	Fish	Hilbert agents
V	11.31 ± 0.18	11.91 ± 0.14
r_w	4.11 ± 0.12	3.90 ± 0.05
d	7.93 ± 0.17	8.27 ± 0.09
P	0.954 ± 0.002	0.957 ± 0.002

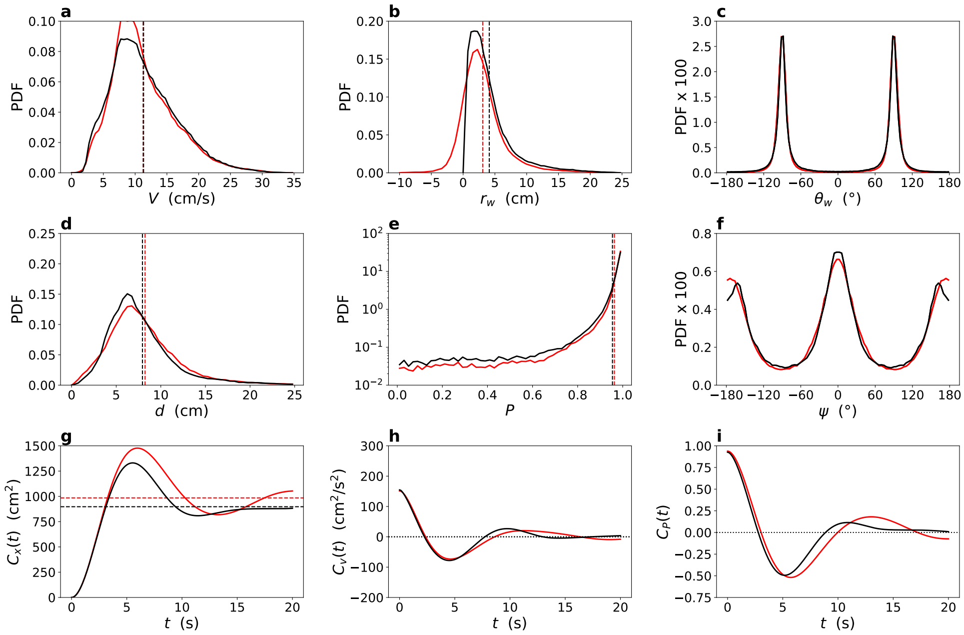
$N = 5$	Fish	Hilbert agents
V	15.26 ± 0.13	15.29 ± 0.10
r_w	7.08 ± 0.11	7.17 ± 0.05
d	4.06 ± 0.04	5.15 ± 0.05
P	0.964 ± 0.001	0.943 ± 0.002
R_{Gyr}	7.62± 0.07	10.42 ± 0.08

Extended Data Table 1: We report the mean and standard error for the PDF of the observables appearing in Fig. 3 (for $N = 2$ individuals) and Fig. 4 (for $N = 5$ individuals). The speed V is expressed in cm/s, while the distance to the wall, r_w, the distance between two nearest neighbors, d, and the gyration radius, R_{Gyr} (only for $N = 5$), are expressed in cm. Finally, the polarization, P, is without unit and between 0 and 1.

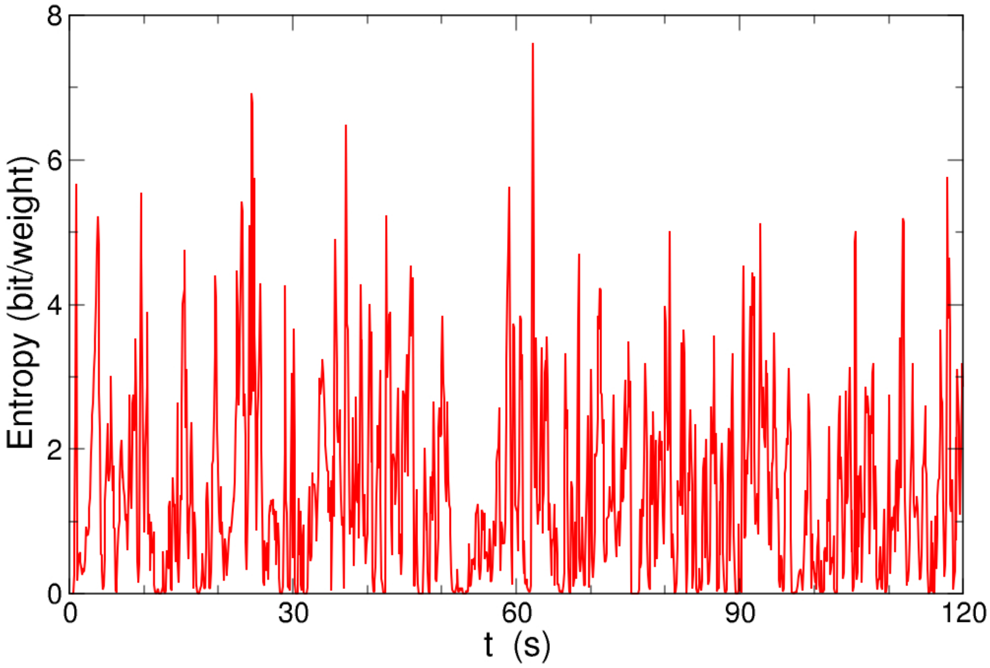
Extended Data Figures



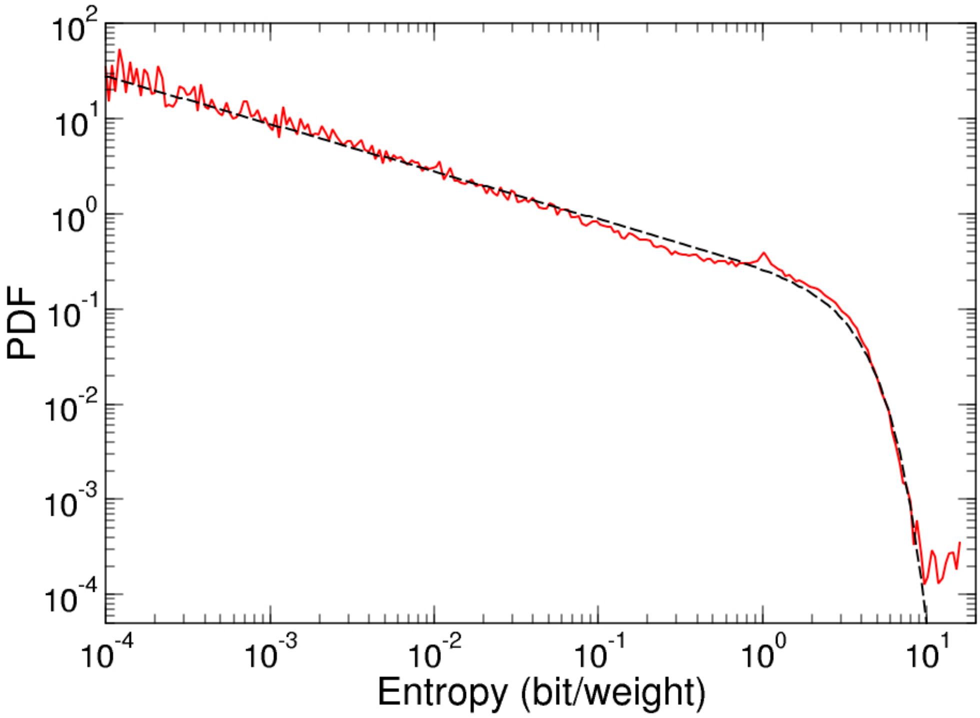
Extended Data Fig. 1: Classification using the Hilbert kernel: A simple example is shown, with two classes of points drawn from a mixture of 2D unit Normal distributions, with mean separated by 2. The points are shown in green and red colors (1000 points of each class). The red vertical line is the Bayes classification boundary. The yellow and blue colored regions are the Hilbert-predicted classification regions for the green and red points. The islands of blue in yellow (and vice versa) are due to the interpolative nature of the classifier, and correspond to the phenomenon of adversarial examples which are guaranteed for interpolating classifiers on noisy data.



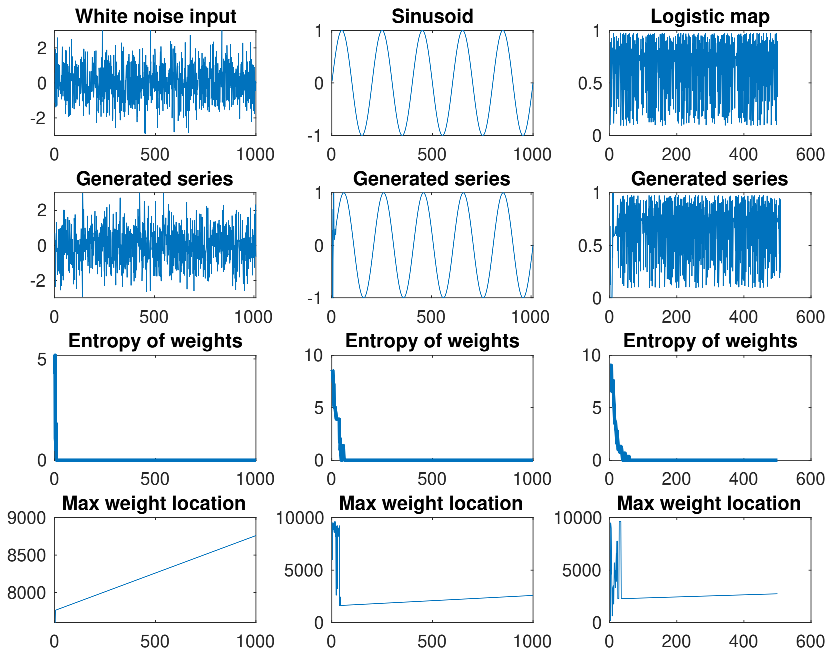
Extended Data Fig. 2: **Behavior of 2 Hilbert fish without the tank wall.** This figure is the analog of Fig. 3 in the main text (also for a memory $M = 2$), but in the case where the presence of the tank wall is not enforced in the Hilbert model. The different panels show the 9 observables used to characterize the individual (**a-c**) and collective (**d-f**) behavior, and the time correlations in the system (**g-i**): **a**, PDF of the speed, V; **b**, PDF of the distance to the wall, r_{w}; **c**, PDF of the heading angle relative to the normal to the wall, θ_{w}; **d**, PDF of the distance between the pair of individuals, d; **e**, PDF of the group polarization, $P = \left|\cos(\Delta\phi/2)\right|$, where $\Delta\phi$ is the relative heading angle; **f**, PDF of the viewing angle at which an individual perceives the other individual, ψ. See Fig. 2a and b in the main text for a visual representation of the main variables. **g**, Mean squared displacement, $C_x(t)$, and its asymptotic limit, $C_x(\infty) = 2\langle r^2 \rangle$ (dotted lines); **h**, Velocity autocorrelation, $C_v(t)$; **i**, Polarization autocorrelation, $C_P(t)$. The black PDFs correspond to experiments, while the red PDFs correspond to the predictions of the Hilbert generative model. The plots are on the same scale as in Fig. 3 in the main text, except for r_{w}, for which the horizontal axis has been extended to negative values of r_{w} corresponding to instances where an individual is observed outside the limits of the experimental circular tank. Yet, the Hilbert fish spend 87 % of the time strictly within the tank limits, and when they wander outside the tank, their average excursion distance from the wall is only 1.3 cm. These excursions are responsible for the upward and rightward shift of the peak of $C_x(t)$ and for the larger asymptotic limit, $C_x(\infty) = 2\langle r^2 \rangle \approx 980\,\mathrm{cm}^2$ (compared to $C_x(\infty) \approx 900\,\mathrm{cm}^2$ for fish or for the Hilbert model implementing the rejection procedure enforcing the presence of the tank wall).



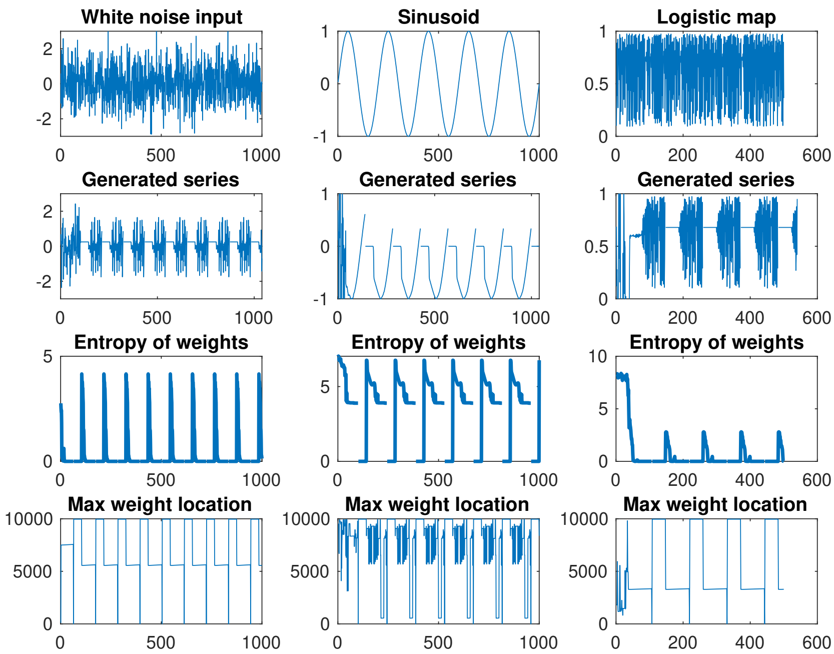
Extended Data Fig. 3: **Entropy time series for 2 Hilbert fish.** We plot a 2-minute time series of the entropy for 2 Hilbert fish, and for a memory $M = 2$. The entropy S can be interpreted as $\log_2 \mathcal{N}$, where $\mathcal{N}$ is the effective number of real fish configurations used to predict the acceleration of the Hilbert fish. The time series exhibits short periods where $\mathcal{N} \approx 1$ ($S \approx 0$), when the Hilbert scheme has essentially selected a unique real fish configuration ("copying"). This short time series also presents three short periods when $\mathcal{N} > 64$ ($S > 6$). The PDF of the entropy computed over much longer time is shown in Extended Data Fig. 4.



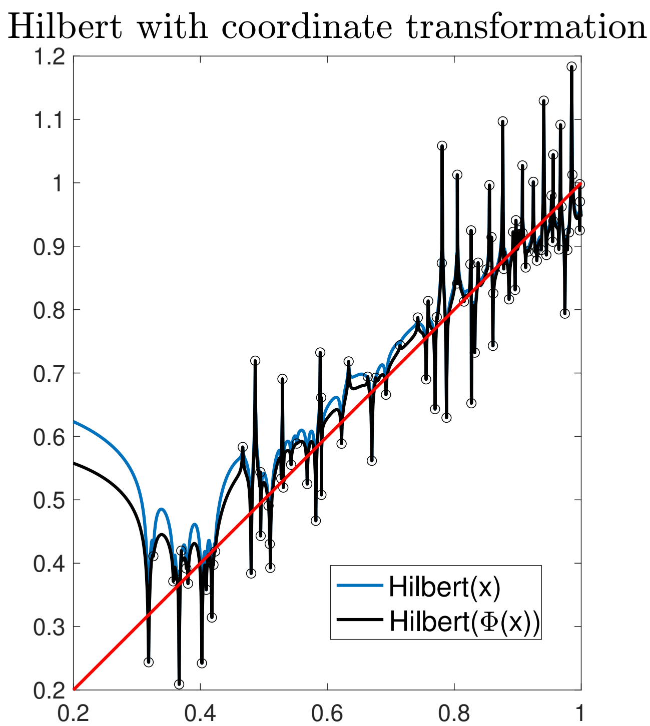
Extended Data Fig. 4: **Entropy distribution for 2 Hilbert fish.** We plot the PDF of the entropy for 2 Hilbert fish (for a memory $M = 2$) resulting from an effective simulation time of 3 hours. The PDF of the entropy is reasonably well fitted by the normalized functional form $\rho(S) = (2\,\Gamma[5/4])^{-1}\,(S_c\,S)^{-1/2}\,\exp\left(-(S/S_c)^2\right)$, although the fit does not capture some outliers at $S > 10$. The fitted cut-off entropy scale, $S_c \approx 3.61$, corresponds to $\mathcal{N}_c = 2^{S_c} \approx 12.2$ real fish configurations contributing to the acceleration prediction, while the mean entropy $\langle S\rangle \approx 1.37$ corresponds to $2^{\langle \log_2 \mathcal{N}\rangle} = 2^{\langle S\rangle} \approx 2.6$ configurations. The mean number of configurations used for a prediction is $\langle \mathcal{N}\rangle = \langle 2^S\rangle$, and is dominated by outliers. If this average is restrained to instance where $S \leq 10$, one finds $\langle \mathcal{N}\rangle_{S\leq 10} \approx 5.2$ (our fit $\rho(S)$ would predict $\langle \mathcal{N}\rangle_{S\leq 10} \approx 4.7$), whereas the average including all data is $\langle \mathcal{N}\rangle \approx 35$. Also note the small peak in the PDF near $S = 1$, corresponding to $\mathcal{N} = 2$ relevant configurations contributing almost equally to the Hilbert prediction. Yet, during the simulation, entropies as high as $S \sim 15$ were recorded, corresponding to $\mathcal{N} \sim 32768$ fish configurations effectively considered by the Hilbert kernel. Compared to kNN methods, the Hilbert interpolation scheme is hence able to adapt the effective number of used data for the prediction to the properties of the input vector. See also Extended Data Fig. 3 for a short time series of the entropy.



Extended Data Fig. 5: Autoregressive generative modeling of time series data: three examples are shown of signals generated by Eq. (9) with simple training data consisting of N=10000 samples of a single training signal. The three columns respectively show results corresponding to a training signal generated by white Gaussian noise, a sinusoid, and iterates of a logistic regression equation $x_{n+1} = \lambda x_n(1 - x_n)$ for $\lambda = 3.9$. A lag window size of $T = 10$ is used (see the next figure for a lag window size of $T = 40$, and the signal generation is initiated by random initial conditions consisting of T samples of a standard normal distributed variable. The second row shows the generated signal. The third row shows the entropy of the generative weights as a function of generation time, and the fourth row shows the position of the maximum weight in the training signal, also as a function of generation time. Note the "copying" behavior, where the generated signal starts some fragment of the training signal after an initial transient. During the "copying" phase, the weight entropy falls to zero, and the index of the maximum weight increments linearly with time.



Extended Data Fig. 6: Autoregressive generative modeling of time series data: the same signal examples as in the previous extended data figure (Extended Data Fig. 5) are shown, this time with a lag window $T = 40$. In this case, periodic behavior is observed after an initial transient. The periods themselves contain short episodes of "copying" where the entropy falls to zero, and also show short episodes of close to fixed-point behavior of the dynamics, where the generated signal has an almost constant value. The extent of the "copying", periodic or constant behaviors depends on the initial conditions as well as on d.



Extended Data Fig. 7: Impact of an initial coordinate transformation $x \to \Phi(x)$. This example shows two Hilbert kernel fits, with the black curve corresponding to the original kernel, and the blue curve corresponding to the generalized kernel corresponding to replacing $x \to \Phi(x)$ in the weights. The samples x_i ($i = 1..100$) are chosen so that the transformed coordinates $\Phi(x_i)$ have a uniform distribution. The red line corresponds to $y = x$ and uncorrelated Gaussian noise with $\sigma = 0.1$ is added to produce the noisy samples. The two regression functions both interpolate, but show slight differences, especially in the data-sparse region, with the transformed weights (that produce uniform sampling of x) being a bit closer to the noise-free function. As proven in the paper, both estimates are statistically consistent and have the same large-sample asymptotics in the leading order, but the sub-leading order behavior will generally depend on Φ in conjunction with the other details of the problem.

Appendix A Proofs of the theorems

A.1 Preliminaries

In the following, $x \in \Omega^\circ$ so that $\rho(x) > 0$, and we will assume for simplicity that the distribution ρ is continuous at x.

For the proof of our results, we will often exploit the following integral relation, valid for $\beta > 0$ and $z > 0$,

$$\frac{1}{\Gamma(\beta)} \int_0^{+\infty} t^{\beta-1} \mathrm{e}^{-t\,z}\, dt = z^{-\beta}. \tag{A1}$$

In addition, we define

$$\psi(x,t) \coloneqq \int \rho(x+y) \mathrm{e}^{-\frac{t}{||y||^d}}\, d^d y, \tag{A2}$$

which will play a central role. We note that $\psi(x,0) = 1$, and that $t \mapsto \psi(x,t)$ is a continuous and strictly decreasing function of t. It is even infinitely differentiable at any $t > 0$, but not necessarily at $t = 0$. In fact, for a fixed x, controlling the behavior of $1 - \psi(x,t)$ when $t \to 0$ will be essential to obtain our results.

A.2 Moments of the weights: large n behavior

In this section, we provide a complete proof of Theorem 1. Several other theorems will use the same method of proof, and some basic steps will not be repeated in their proof.

Using Eq. (A1) for $\beta > 0$, we can express powers of the weight function as

$$w_0^\beta(x) = \frac{1}{||x-x_0||^{\beta d}} \frac{1}{\Gamma(\beta)} \int_0^{+\infty} t^{\beta-1} \mathrm{e}^{-t\,||x-x_0||^{-d} - t\sum_{i=1}^n ||x-x_i||^{-d}}\, dt. \tag{A3}$$

By taking the expected value over the $n+1$ independent random variables X_i, we obtain

$$\mathbb{E}\left[w_0^\beta(x)\right] = \frac{1}{\Gamma(\beta)} \int_0^{+\infty} t^{\beta-1} \psi^n(x,t) \phi_\beta(x,t)\, dt, \tag{A4}$$

with

$$\phi_\beta(x,t) \coloneqq \int \rho(x+y) \frac{\mathrm{e}^{-\frac{t}{||y||^d}}}{||y||^{\beta d}}\, d^d y, \tag{A5}$$

which is also a strictly decreasing function of t, continuous at any $t > 0$ (in fact, infinitely differentiable for $t > 0$).

Note that the exchange of the integral over t and over $\vec{x} = (x_0, x_1, ..., x_n)$ used to obtain Eq. (A4) is justified by the Fubini theorem, by first noting that the function $\vec{x} \mapsto w_0^\beta(x) \prod_{i=0}^n \rho(x_i)$ is in $L^1(\mathbb{R}^d)$, since $0 \le w_0^\beta(x) \le 1$, and since ρ is obviously in $L^1(\mathbb{R}^d)$. Moreover, the function $t \mapsto t^{\beta-1}\psi^n(x,t)\phi_\beta(x,t) > 0$ is also in $L^1(\mathbb{R})$. Indeed, we will show below that it decays fast enough when $t \to +\infty$ (see Eqs. (A7-A16)), ensuring the convergence of its integral at $+\infty$, and that it is bounded (and continuous) near $t = 0$ (see Eqs. (A29-A34)), ensuring that this function is integrable at $t = 0$.

For $\beta = 1$, $\phi_1 = -\partial_t \psi$, and we obtain $\mathbb{E}\left[w_0(x)\right] = \frac{1}{n+1}$, as expected. In the following, we first focus on the case $\beta > 1$, before addressing the cases $0 < \beta < 1$ and $\beta < 0$ at the very end of this section.

We now introduce t_1 and t_2 (to be further constrained later) such that $0 < t_1 < t_2$. We then express the integral of Eq. (A4) as the sum of corresponding integrals $I_1 + I_{12} + I_2$. I_1 is the integral between 0 and t_1, I_{12} the integral between t_1 and t_2, and I_2 the integral between t_2 and $+\infty$. Thus, we have

$$I_1 \le \mathbb{E}\left[w_0^\beta(x)\right] \le I_1 + I_{12} + I_2, \tag{A6}$$

provided these integral exists, which we will show below, by providing upper bounds for I_2 and I_{12}, and tight lower and upper bound for the leading term I_1.

Bound for I_2

For any $R \ge 1$, we can write the integral defining $\psi(x,t)$

$$\begin{aligned}\psi(x,t) &= \int_{||y||\le R} + \int_{||y||\ge R} &\text{(A7)}\\ &\le \mathrm{e}^{-\frac{t}{R^d}} + \int_{||y||\ge R} \rho(x+y)\frac{||y||^2}{R^2}\, d^d y, &\text{(A8)}\\ &\le \mathrm{e}^{-\frac{t}{R^d}} + \frac{C_x}{R^2}, &\text{(A9)}\end{aligned}$$

with $C_x = \sigma_\rho^2 + ||x-\mu_\rho||^2$ depending on the mean μ_ρ and variance σ_ρ^2 of the distribution ρ. Similarly, for $\phi_\beta(x,t)$, we obtain the bound

$$\phi_\beta(x,t) \le \frac{1}{R^{\beta d}}\mathrm{e}^{-\frac{t}{R^d}} + \frac{C_x}{R^{2+\beta d}}, \tag{A10}$$

valid for $t \ge \max(1,\beta)$ and $R \le r_t$, where $r_t = (t/\beta)^{1/d} \ge 1$ is the location of the maximum of the function $r \mapsto \frac{\mathrm{e}^{-\frac{t}{r^d}}}{r^{\beta d}}$.

We now set $R = t^{\frac{s}{d}}$, with $0 < s < 1$, and take $T_2' \ge \max(1, \beta, \beta^{1/(1-s)})$ (so that $1 \le R \le r_t$) is large enough such that the following conditions are satisfied for $t \ge t_2 \ge T_2'$,

$$\mathrm{e}^{-\frac{t}{R^d}} = \mathrm{e}^{-t^{1-s}} \le \frac{C_x}{t^{\frac{2s}{d}}}, \tag{A11}$$

$$\frac{1}{R^{\beta d}}\mathrm{e}^{-\frac{t}{R^d}} = \frac{1}{t^{\beta s}}\mathrm{e}^{-t^{1-s}} \le \frac{C_x}{t^{\frac{2s}{d}+\beta s}}. \tag{A12}$$

Hence, for $t \ge t_2 \ge T_2'$, we obtain

$$\psi(x,t) \le \frac{2C_x}{t^{\frac{2s}{d}}}, \tag{A13}$$

$$\phi_\beta(x,t) \le \frac{2C_x}{t^{\frac{2s}{d}+\beta s}}. \tag{A14}$$

In addition, we also impose $t_2 \ge T_2'' = (4C_x)^{d/(2s)}$, so that $\frac{2C_x}{t^{\frac{2s}{d}}} \le \frac{1}{2}$, for any $t \ge T_2 = \max(T_2', T_2'')$. We can now exploit the resulting bounds for $\psi(x,t)$ and $\phi_\beta(x,t)$ in Eq. (A13) and Eq. (A14) to compute an explicit bound for I_2, for any given $t_2 \ge T_2$:

$$I_2 = \frac{1}{\Gamma(\beta)}\int_{t_2}^{+\infty} t^{\beta-1}\psi^n(x,t)\phi_\beta(x,t)\,dt \le \frac{1}{\Gamma(\beta)}\int_{t_2}^{+\infty} t^{\beta(1-s)-1}\left(\frac{2C_x}{t^{\frac{2s}{d}}}\right)^{n+1} dt. \tag{A15}$$

The integral in the right-hand side of Eq. (A15) only converges for $s > \frac{1}{1+\frac{2(n+1)}{\beta d}}$ (remember that we also impose $s < 1$), and we then set $s = \frac{1}{1+\frac{2}{\beta d}}$, which ensures its convergence for any $n \ge 1$. Performing this integral and using the fact that $\frac{2C_x}{t_2^{\frac{2s}{d}}} \le \frac{1}{2}$, we finally obtain

$$I_2 \le C_x \frac{d+\frac{2}{\beta}}{\Gamma(\beta)} \times \frac{1}{n\,2^n}. \tag{A16}$$

We hence obtain the convergence of I_2, which, along with the bounds for I_1 and I_{12} below, justifies our use of Fubini theorem to obtain Eq. (A4). Note that the above bound essentially decays exponentially with n, under the stated conditions.

Bound for I_{12}

Again, exploiting the fact that $\psi(x,t)$ and $\phi_\beta(x,t)$ are strictly decreasing functions of t, we obtain

$$I_{12} \le \frac{\phi_\beta(x,t_1)t_2^\beta}{\Gamma(\beta)} \times \psi^n(x,t_1), \tag{A17}$$

where we note that $\psi(x,t_1) < 1$, for any $t_1 > 0$, implying that this bound decays exponentially with n.

Bound for I_1

We first want to obtain bounds for $1-\psi(x,t)$, where $0 \le t \le t_1$, with $t_1 > 0$ to be constrained below. In addition, exploiting the continuity of ρ at x and the fact that $\rho(x) > 0$, we introduce ε satisfying $0 < \varepsilon < 1/4$, and define $\lambda > 0$ small enough so that the ball $B(x,\delta) \subset \Omega^\circ$, and $||y|| \le \lambda \implies |\rho(x+y)-\rho(x)| \le \varepsilon\rho(x)$. Exploiting this definition, we obtain the following lower and upper bounds

$$1-\psi(x,t) \ge (1-\varepsilon)\rho(x)\int_{||y||\le\lambda}\left(1-\mathrm{e}^{-\frac{t}{||y||^d}}\right)d^dy, \tag{A18}$$

$$1-\psi(x,t) \;\leq\; (1+\varepsilon)\rho(x)\int_{||y||\leq\lambda}\left(1-\mathrm{e}^{-\frac{t}{||y||^d}}\right)\,d^dy \tag{A19}$$

$$+\int_{||y||\geq\lambda}\rho(x+y)\left(1-\mathrm{e}^{-\frac{t}{\lambda^d}}\right)\,d^dy, \tag{A20}$$

$$\leq\;(1+\varepsilon)\rho(x)\int_{||y||\leq\lambda}\left(1-\mathrm{e}^{-\frac{t}{||y||^d}}\right)\,d^dy+\frac{t}{\lambda^d}. \tag{A21}$$

The integral appearing in these bounds can be simplified by using radial coordinates:

$$\int_{||y||\leq\lambda}\left(1-\mathrm{e}^{-\frac{t}{||y||^d}}\right)\,d^dy,\;=\;S_d\int_0^\lambda\left(1-\mathrm{e}^{-\frac{t}{r^d}}\right)r^{d-1}\,dr, \tag{A22}$$

$$=\;V_d t\int_{\frac{t}{\lambda^d}}^{+\infty}\frac{1-\mathrm{e}^{-u}}{u^2}\,du, \tag{A23}$$

where S_d and $V_d=\frac{S_d}{d}$ are respectively the surface and the volume of the d-dimensional unit sphere, and we have used the change of variable $u=\frac{t}{r^d}$.

We note that for $0<z\leq 1$, we have

$$\int_z^{+\infty}\frac{1-\mathrm{e}^{-u}}{u^2}\,du=-\ln(z)+\int_z^1\frac{1-u-\mathrm{e}^{-u}}{u^2}\,du+\int_1^{+\infty}\frac{1-\mathrm{e}^{-u}}{u^2}\,du. \tag{A24}$$

Exploiting this result and now imposing $t_1\leq\lambda^d$, we have, for any $t\leq t_1$

$$\ln\left(\frac{C_-}{t}\right)\;\leq\;\int_{\frac{t}{\lambda^d}}^{+\infty}\frac{1-\mathrm{e}^{-u}}{u^2}\,du\quad\leq\quad\ln\left(\frac{C_+}{t}\right), \tag{A25}$$

$$\ln(C_+)\;=\;d\ln(\lambda)+\int_1^{+\infty}\frac{1-\mathrm{e}^{-u}}{u^2}\,du, \tag{A26}$$

$$\ln(C_-)\;=\;\ln(C_+)+\int_0^1\frac{1-u-\mathrm{e}^{-u}}{u^2}\,du. \tag{A27}$$

Combining these bounds with Eq. (A18) and Eq. (A21), we have shown the existence of two x-dependent constants $D_\pm$ such that, for $0\leq t\leq t_1\leq\lambda^d$, we have

$$(1-\varepsilon)V_d\rho(x)\,t\ln\left(\frac{D_-}{t}\right)\leq 1-\psi(x,t)\leq(1+\varepsilon)V_d\rho(x)\,t\ln\left(\frac{D_+}{t}\right). \tag{A28}$$

In addition, we will also choose $t_1<D_\pm/3$, such that the two functions $t\ln\left(\frac{D_\pm}{t}\right)$ are positive and strictly increasing for $0\leq t\leq t_1$. t_1 is also taken small enough such that the two bounds in Eq. (A28) are always less than $1/2$, for $0\leq t\leq t_1$ (both bounds vanish when $t\to 0$).

We now obtain efficient bounds for $\phi_\beta(x,t)$, for $0 \le t \le t_1$. Proceeding similarly as above, we obtain

$$\phi_\beta(x,t) \ \geq\ (1-\varepsilon)\rho(x) \int_{||y||\le\lambda} \frac{\mathrm{e}^{-\frac{t}{||y||^d}}}{||y||^{\beta d}}\, d^d y, \tag{A29}$$

$$\phi_\beta(x,t) \ \leq\ (1+\varepsilon)\rho(x) \int_{||y||\le\lambda} \frac{\mathrm{e}^{-\frac{t}{||y||^d}}}{||y||^{\beta d}}\, d^d y + \frac{1}{\lambda^{\beta d}}. \tag{A30}$$

Again, the integral appearing in these bounds can be rewritten as

$$\int_{||y||\le\lambda} \frac{\mathrm{e}^{-\frac{t}{||y||^d}}}{||y||^{\beta d}}\, d^d y = S_d \int_0^\lambda r^{d(1-\beta)-1} \mathrm{e}^{-\frac{t}{r^d}}\, dr. \tag{A31}$$

For $0<\beta<1$, the integral of Eq. (A31) is finite for $t=0$, ensuring the existence of $\phi_\beta(x,0)$ and the fact that $t \mapsto t^{\beta-1}\psi(x,t)\phi_\beta(x,t)$ belongs to $L^1(\mathbb{R})$ (hence, justifying our use of Fubini theorem for $0<\beta<1$). For $\beta>1$, we have

$$\int_{||y||\le\lambda} \frac{\mathrm{e}^{-\frac{t}{||y||^d}}}{||y||^{\beta d}}\, d^d y \ = \ V_d\, t^{1-\beta} \int_{\frac{t}{\lambda^d}}^{+\infty} u^{\beta-2} \mathrm{e}^{-u}\, du. \tag{A32}$$

$$\underset{t\to 0}{\sim}\ V_d \Gamma(\beta-1) t^{1-\beta}. \tag{A33}$$

This integral diverges when $t \to 0$ and the constant term $\lambda^{-\beta d}$ in Eq. (A30) can be made as small as necessary (by a factor less than ε) compared to this leading integral term, for a small enough t_1. Similarly, we can choose t_1 small enough so that the integral Eq. (A31) is approached by the asymptotic result of Eq. (A33) up to a factor ε. Thus, we find that for $0 \le t \le t_1$, one has

$$(1-2\varepsilon)V_d\rho(x)\Gamma(\beta-1)t^{1-\beta} \le \phi_\beta(x,t) \le (1+3\varepsilon)V_d\rho(x)\Gamma(\beta-1)t^{1-\beta}. \tag{A34}$$

This shows that $t^{\beta-1}\phi_\beta(x,t)$ has a smooth limit equal to $V_d\rho(x)\Gamma(\beta-1)$, when $t \to 0$, so that, combined with the finite upper bound for I_2, $t \mapsto t^{\beta-1}\psi(x,t)\phi_\beta(x,t)$ belongs to $L^1(\mathbb{R})$, for $\beta>1$, and hence for all $\beta>0$. Hence, the use of the Fubini theorem to derive Eq. (A4) has been justified.

Now combining the bounds for $\psi(x,t)$ and $\phi_\beta(x,t)$, we obtain

$$I_1 \ \geq\ (1-2\varepsilon)\frac{1}{\beta-1}V_d\rho(x) \int_0^{t_1} \left(1-(1+\varepsilon)V_d\rho(x)\, t \ln\left(\frac{D_+}{t}\right)\right)^n dt, \tag{A35}$$

$$I_1 \ \leq\ (1+3\varepsilon)\frac{1}{\beta-1}V_d\rho(x) \int_0^{t_1} \left(1-(1-\varepsilon)V_d\rho(x)\, t \ln\left(\frac{D_-}{t}\right)\right)^n dt. \tag{A36}$$

Asymptotic behavior of I_1 *and* $\mathbb{E}\left[w_0^\beta(x)\right]$

We will show below that

$$\int_0^{t_1} \left(1 - E_\pm t \ln\left(\frac{D_\pm}{t}\right)\right)^n dt \underset{n\to+\infty}{\sim} \frac{1}{E_\pm n \ln(n)}, \tag{A37}$$

where $E_\pm = (1 \mp \varepsilon)V_d\rho(x)$. For a given x, and for t_1 and t_2 satisfying the requirements mentioned above, the upper bounds for I_{12} (see Eq. (A17)) and I_2 (see Eq. (A16)) appearing in Eq. (A6) both decay exponentially with n and can hence be made arbitrarily small compared to I_1 which decays as $1/(n\,\ln(n))$.

Finally, assuming for now the result of Eq. (A37) (to be proven below), we have obtained the exact asymptotic result

$$\mathbb{E}\left[w_0^\beta(x)\right] \underset{n\to+\infty}{\sim} \frac{1}{(\beta-1)n\ln(n)}. \tag{A38}$$

Proof of Eq. (A37)

We are then left to prove the result of Eq. (A37). First, we will use the fact that, for $0 \le z \le z_1 < 1$, one has

$$\mathrm{e}^{-\mu z} \le 1 - z \le \mathrm{e}^{-z}, \tag{A39}$$

where $\mu = -\ln(1-z_1)/z_1$. We can apply this result to the integral of Eq. (A37), using $z_1^\pm = E_\pm t_1 \ln(D_\pm/t_1) > 0$. Note that $0 < t_1 < D_\pm/3$ and hence $z_1^\pm > 0$ can be made as close to 0 as desired, and the corresponding $\mu_\pm > 1$ can be made as close to 1 as desired. Thus, in order to prove Eq. (A37), we need to prove the following equivalent

$$I_n = \int_0^{t_1} \mathrm{e}^{-nEt\ln\left(\frac{D}{t}\right)}\, dt \underset{n\to+\infty}{\sim} \frac{1}{En\ln(n)}, \tag{A40}$$

for an integral of the form appearing in Eq. (A40). Let us mention again that t_1 has been taken small enough so that the function $t \mapsto t\ln\left(\frac{D}{t}\right)$ is positive and strictly increasing (with its maximum at $t_{\max} = D/e < t_1$), for $0 \le t \le t_1$.

We now take n large enough so that $\frac{\ln(n)}{n} < t_1$ and $E\ln(n) > 1$. One can then write

$$I_n = \frac{1}{n}\int_0^{\ln(n)} \mathrm{e}^{-Eu\ln\left(\frac{Dn}{u}\right)}\, du + \int_{\frac{\ln(n)}{n}}^{t_1} \mathrm{e}^{-nEt\ln\left(\frac{D}{t}\right)}\, dt = J_n + K_n, \tag{A41}$$

$$J_n \le \frac{1}{n}\int_0^{1/E} \mathrm{e}^{-Eu\ln(DEn)}\, du + \frac{1}{n}\int_{1/E}^{\ln(n)} \mathrm{e}^{-Eu\ln\left(\frac{Dn}{\ln(n)}\right)}\, du, \tag{A42}$$

$$\le \frac{1}{E\,n\ln(D\,En)} + \frac{\ln(n)}{D\,E\,n^2\ln\left(\frac{Dn}{\ln(n)}\right)}, \tag{A43}$$

$$K_n \le \int_{\frac{\ln(n)}{n}}^{+\infty} \mathrm{e}^{-nEt\ln\left(\frac{D}{t_1}\right)}\, dt \le \frac{1}{E\,n^{1+E\ln\left(\frac{D}{t_1}\right)}\ln\left(\frac{D}{t_1}\right)}. \tag{A44}$$

When $n \to +\infty$, we hence find that the upper bound I_n^+ of I_n satisfies

$$I_n^+ \underset{n\to+\infty}{\sim} \frac{1}{E\, n \ln (DEn)} \underset{n\to+\infty}{\sim} \frac{1}{E\, n \ln (n)}. \tag{A45}$$

Let us now prove a similar result for a lower bound of I_n by considering n large enough so that $nEt_1 > 1$, and by introducing δ satisfying $0 \le \delta < 1/e$:

$$I_n = \frac{1}{nE} \int_0^{nEt_1} \mathrm{e}^{-u \ln(DEn) + u \ln(u)}\, du, \tag{A46}$$

$$\ge \frac{1}{nE} \int_0^{\delta} \mathrm{e}^{-u \ln(DEn) + \delta \ln(\delta)}\, du, \tag{A47}$$

$$\ge \frac{\mathrm{e}^{\delta \ln(\delta)}}{nE \ln (DEn)} \left(1 - (DEn)^{-\delta}\right) = I_n^-(\delta). \tag{A48}$$

Hence, for any $0 \le \delta < 1/e$ which can be made arbitrarily small, and for n large enough, we find that $I_n \ge I_n^-(\delta)$, with

$$I_n^-(\delta) \sim \frac{\mathrm{e}^{\delta \ln(\delta)}}{E\, n \ln (DEn)} \sim \frac{\mathrm{e}^{\delta \ln(\delta)}}{E\, n \ln (n)}. \tag{A49}$$

Eq. (A49) combined with the corresponding result of Eq. (A45) for the upper bound I_n^+ finally proves Eq. (A40), and ultimately, Eq. (A38) and Theorem 1 for the asymptotic behavior of the moment $\mathbb{E}\left[w_0^\beta(x)\right]$, for $\beta > 1$.

Entropy (moment for "$\beta = 1^-$")

We define the information entropy, $S(x)$, by

$$S(x) = -\sum_{i=0}^{n} w_i(x) \log[w_i(x)]. \tag{A50}$$

If the weights are equidistributed over $\mathcal{N}$ data, one obtains $S = -\mathcal{N} \times 1/\mathcal{N} \log(1/\mathcal{N}) = \log(\mathcal{N})$, and $\mathrm{e}^S = \mathcal{N}$ indeed represents the number of contributing data. The expectation value of the entropy reads

$$\mathbb{E}[S(x)] = -(n+1)\, \mathbb{E}[w_0(x) \ln(w_0(x))]. \tag{A51}$$

In order to evaluate Eq. (A51), we use an integral representation in the spirit of Eq. (A1), valid for any $z > 0$,

$$\int_0^{+\infty} (\ln(t) + \gamma)\, \mathrm{e}^{-t\, z}\, dt = -\frac{\ln(z)}{z}, \tag{A52}$$

where γ is Euler's constant. Using Eq. (A52), we find

$$-w_0(x)\ln(w_0(x)) = -\frac{1}{||x-x_0||^{-d}}\int_0^{+\infty} \mathrm{e}^{-t\,||x-x_0||^{-d}-t\,\sum_{i=1}^{n}||x-x_i||^{-d}} \times \Big(\ln(||x-x_0||^{-d})+\ln(t)+\gamma\Big)\,dt. \tag{A53}$$

By taking the expected value over the $n+1$ independent random variables X_i, we obtain

$$\mathbb{E}\left[-w_0(x)\ln(w_0(x))\right] = -\int_0^{+\infty}\psi^n(x,t)\left(\Phi_1(x,t)+(\ln(t)+\gamma)\phi_1(x,t)\right)\,dt, \tag{A54}$$

with

$$\Phi_1(x,t) := \int \rho(x+y)\,\mathrm{e}^{-\frac{t}{||y||^d}}\,\frac{\ln\left(||y||^{-d}\right)}{||y||^d}\,d^dy, \tag{A55}$$

which is continuous at any $t>0$ (in fact, infinitely differentiable for $t>0$). In addition, $\phi_1(x,t) = -\partial_t\psi(x,t)$ has been defined in Eq. (A5).

By exploiting the same method used to bound $\phi_\beta(x,t)$ (see Eq. (A34) and above it), we find that

$$\Phi_1(x,t) \underset{t\to 0}{\sim} \frac{1}{2}V_d\rho(x)\ln^2(t), \tag{A56}$$

$$\phi_1(x,t) \underset{t\to 0}{\sim} -V_d\rho(x)\ln(t), \tag{A57}$$

where Eq. (A57) is fully consistent with Eq. (A28) (by naively differentiating Eq. (A28)).

Finally, exploiting Eqs. (A56,A57), the integral of Eq. (A54) can be evaluated with the same method as in the previous section, leading to

$$\mathbb{E}\left[-w_0(x)\ln(w_0(x)\right] \underset{n\to+\infty}{\sim} \frac{1}{2}V_d\rho(x)\int_0^{t_1}\mathrm{e}^{-nV_d\rho(x)t\ln\left(\frac{D_\pm}{t}\right)}\ln^2(t)\,dt, \tag{A58}$$

$$\underset{n\to+\infty}{\sim} \frac{1}{2}\frac{\ln(n)}{n}. \tag{A59}$$

This last result proves the second part of Theorem 1 (see also the heuristic discussion below Theorem 1) for the expected value of the entropy:

$$\mathbb{E}[S(x)] = -(n+1)\,\mathbb{E}[w_0(x)\ln(w_0(x))] \underset{n\to+\infty}{\sim} \frac{1}{2}\ln(n). \tag{A60}$$

Moments of order $0<\beta<1$

The integral representation Eq. (A1) allows us to also explore moments of order $0<\beta<1$. In that case $\kappa_\beta(x) = \phi_\beta(x,0) < \infty$ is finite, with

$$\kappa_\beta(x) = \int \frac{\rho(x+y)}{||y||^{\beta d}}\,d^dy. \tag{A61}$$

By retracing the different steps of our proof in the case $\beta > 1$, it is straightforward to show that

$$\mathbb{E}\left[w_0^\beta(x)\right] \underset{n\to+\infty}{\sim} \frac{\kappa_\beta(x)}{\Gamma(\beta)} \int_0^{t_1} t^{\beta-1} \mathrm{e}^{-nV_d\rho(x)t\ln\left(\frac{D_\pm}{t}\right)}\, dt, \tag{A62}$$

$$\underset{n\to+\infty}{\sim} \frac{\kappa_\beta(x)}{(V_d\rho(x)n\ln(n))^\beta}, \tag{A63}$$

where the equivalent for the integral can be obtained by exploiting the very same method used in our proof of Eq. (A37) above, hence proving the third part of Theorem 1.

We observe that contrary to the universal result of Eq. (A38) for β, the asymptotic equivalent for the moment of order $0 < \beta < 1$ is non-universal and explicitly depends on x and the distribution ρ.

Moments of order $\beta < 0$

Finally, moments of order $\beta < 0$ are unfortunately inaccessible to our methods relying on the integral relation Eq. (A1), which imposes $\beta > 0$. However, we can obtain a few rigorous results for these moments (see also the heuristic discussion just after Theorem 1).

Indeed, for $\beta = -1$, we have

$$\frac{1}{w_0(x)} = 1 + \|x - x_0\|^d \sum_{i=1}^{n} \frac{1}{\|x - x_i\|^d}. \tag{A64}$$

But since we have assumed that $\rho(x) > 0$, $\mathbb{E}[\|x - x_i\|^{-d}] = \int \frac{\rho(x+y)}{||y||^d}\, d^dy$ is infinite and moments of order $\beta \leq -1$ are definitely not defined.

As for the moment of order $-1 < \beta < 0$, it can be easily bounded,

$$\mathbb{E}\left[w_0^\beta(x)\right] \leq 1 + n \int \rho(x+y)||y||^{|\beta|d}\, d^dy \int \frac{\rho(x+y)}{||y||^{|\beta|d}}\, d^dy, \tag{A65}$$

and a sufficient condition for its existence is $\kappa_\beta(x) = \int \rho(x+y)||y||^{|\beta|d}\, d^dy < \infty$ (the other integral, equal to $\kappa_{|\beta|}(x)$, is always finite for $|\beta| < 1$), which proves the last part of Theorem 1.

Numerical distribution of the weights

In the main text below Theorem 1, we presented a heuristic argument showing that the results of Theorem 1 and Theorem 2 (for the Lagrange function; that we prove below) were fully consistent with the weight $W = w_0(x)$ having a long-tailed scaling distribution,

$$P_n(W) = \frac{1}{W_n} p\left(\frac{W}{W_n}\right). \tag{A66}$$

The scaling function p was shown to have a universal tail $p(w) \sim w^{-2}$ and the scale W_n was shown to obey the equation $-W_n \ln(W_n) = n^{-1}$. To the leading order for large n, we have $W_n \sim \frac{1}{n\ln(n)}$, and we can solve this equation recursively to find the next order approximation, $W_n \sim \frac{1}{n\ln(n\ln(n))}$. In Fig. 1b in the main text, we present numerical simulations for the scaling distribution p of the variable $w = W/W_n$, for $n = 65536$, using the estimate $W_n \approx \frac{1}{n\ln(n\ln(n))}$. We observe that $p(w)$ is very well approximated by the function $\hat{p}(w) = \frac{1}{(1+w)^2}$, confirming our non-rigorous results.

The data were generated by drawing random values of $r_i^d = ||x - x_i||^d$ using $(n+1)$ *i.i.d.* random variables a_i uniformly distributed in $[0, 1[$, with the relation $r_i = [a_i/(1-a_i)]^{1/d}$, and by computing the resulting weight $W = r_i^{-d}/\sum_{j=0}^{n} r_j^{-d}$. This corresponds to a distribution of $||x - x_i||$ given by $\rho(x - x_i) = 1/V_d/(1 + ||x - x_i||^d)^2$.

A.3 Lagrange function: scaling limit

In this section, we prove Theorem 2 for the scaling limit of the Lagrange function $L_0(x) = \mathbb{E}_{X|x_0}[w_0(x)]$. Exploiting again Eq. (A1), the expected Lagrange function can be written as

$$L_0(x) = \|x - x_0\|^{-d} \int_0^{+\infty} \psi^n(x,t) \mathrm{e}^{-t\|x-x_0\|^{-d}}\, dt, \tag{A67}$$

where $\psi(x,t)$ is again given by Eq. (A2).

For a given $t_1 > 0$, and remembering that $\psi(x,t)$ is a strictly decreasing function of t, with $\psi(x,0) = 1$, we obtain

$$L_1 \le L_0(x) \le L_1 + L_2, \tag{A68}$$

with

$$L_1 = \|x - x_0\|^{-d} \int_0^{t_1} \psi^n(x,t) \mathrm{e}^{-t\|x-x_0\|^{-d}}\, dt, \tag{A69}$$

$$L_2 = \mathrm{e}^{-t_1\|x-x_0\|^{-d}}. \tag{A70}$$

For $\varepsilon > 0$ and a sufficiently small $t_1 > 0$ (see section A.2), we can use the bound for $\psi(x,t)$ obtained in section A.2, to obtain

$$L_1 \ge (1-2\varepsilon)\frac{1}{\|x-x_0\|^d}\int_0^{t_1}\left(1-(1+\varepsilon)V_d\rho(x)\,t\ln\left(\frac{D_+}{t}\right)\right)^n \mathrm{e}^{-\frac{t}{\|x-x_0\|^d}}\, dt \tag{A71}$$

$$L_1 \le (1+3\varepsilon)\frac{1}{\|x-x_0\|^d}\int_0^{t_1}\left(1-(1-\varepsilon)V_d\rho(x)\,t\ln\left(\frac{D_-}{t}\right)\right)^n \mathrm{e}^{-\frac{t}{\|x-x_0\|^d}}\, dt \tag{A72}$$

Then, proceeding exactly as in section A.2, it is straightforward to show that L_1 can be bounded (up to factors $1 + O(\varepsilon)$) by the two integrals $L_1^{\pm}$

$$L_1^{\pm} = \frac{1}{\|x-x_0\|^d}\int_0^{t_1} \mathrm{e}^{-n\,V_d\rho(x)\,t\ln\left(\frac{D_\pm}{t}\right)-\frac{t}{\|x-x_0\|^d}}\, dt. \tag{A73}$$

Like in section A.2, we impose $t_1 < D_\pm/3$, such that the two functions $t\ln\left(\frac{D_\pm}{t}\right)$ are positive and strictly increasing for $0 \le t \le t_1$.

We now introduce the scaling variable $z_x(n, x_0) = V_d \rho(x) \|x - x_0\|^d n \log(n)$, so that

$$L_1^{\pm} = \frac{1}{\|x - x_0\|^d} \int_0^{t_1} \mathrm{e}^{-\frac{t}{\|x-x_0\|^d}\left(1+z\frac{\ln\left(D_\pm/t\right)}{\ln(n)}\right)}\, dt = \int_0^{\frac{t_1}{\|x-x_0\|^d}} \mathrm{e}^{-u\left(1+z\frac{\ln\left(D_\pm\|x-x_0\|^{-d}/u\right)}{\ln(n)}\right)}\, du, \tag{A74}$$

where we have used the shorthand notation $z \equiv z_x(n, x_0)$.

For a given real $Z \geq 0$, we now want to study the (scaling) limit of $L_0(x)$ when $n \to \infty$, $\|x - x_0\|^{-d} \to +\infty$ (i.e., $x_0 \to x$), and such that $z_x(n, x_0) \to Z$, which we will simply denote $\lim_Z L_0(x)$. We note that $\lim_Z L_2 = 0$ (see Eq. (A68) and Eq. (A70)), so that we are left to show that $\lim_Z L_1^{\pm} = \frac{1}{1+Z} = \lim_Z L_0(x)$, which will prove Theorem 2.

Exploiting the fact that $u \ln(u) \geq -1/e$, for $u \geq 0$, we obtain

$$L_1^{\pm} \geq \mathrm{e}^{-\frac{z}{e\ln(n)}} \int_0^{\frac{t_1}{\|x-x_0\|^d}} \mathrm{e}^{-u\left(1+z\frac{\ln\left(D_\pm\|x-x_0\|^{-d}\right)}{\ln(n)}\right)}\, du, \tag{A75}$$

$$\geq \frac{1}{1 + z\frac{\ln\left(D_\pm\|x-x_0\|^{-d}\right)}{\ln(n)}} \mathrm{e}^{-\frac{z}{e\ln(n)}} \left(1 - \mathrm{e}^{-\frac{t_1}{\|x-x_0\|^d}}\right). \tag{A76}$$

Since we have $\lim_Z \frac{\ln\left(D_\pm\|x-x_0\|^{-d}\right)}{\ln(n)} = 1$, $\lim_Z \frac{z}{e\ln(n)} = 0$, and $\lim_Z \frac{t_1}{\|x-x_0\|^d} = +\infty$, we find that $L_1^{\pm}$ is bounded from below by a term for which the $\lim_Z$ is $\frac{1}{1+Z}$, with a relative difference of order $1/\ln(n)$ for finite n.

Since we will ultimately take the $\lim_Z$ and hence the limit $x_0 \to x$, we can impose that the upper limit of the last integral in Eq. (A74) satisfies $\frac{t_1}{\|x-x_0\|^d} > 1$. Let us now consider $K > 0$, such that $K < \frac{t_1}{\|x-x_0\|^d}$. We then obtain,

$$L_1^{\pm} \leq \int_0^{K} \mathrm{e}^{-u\left(1+z\frac{\ln\left(D_\pm\|x-x_0\|^{-d}/K\right)}{\ln(n)}\right)}\, du + \int_K^{+\infty} \mathrm{e}^{-u}\, du, \tag{A77}$$

$$\leq \frac{1}{1 + z\frac{\ln\left(D_\pm\|x-x_0\|^{-d}/K\right)}{\ln(n)}} + \mathrm{e}^{-K}. \tag{A78}$$

We can now take K such that $\ln(K) = \left[\ln\left(\frac{t_1}{\|x-x_0\|^d}\right)\right]^\alpha$, for some fixed α satisfying $0 < \alpha < 1$. It is clear that K satisfies $K < \frac{t_1}{\|x-x_0\|^d}$. In addition, we have $\lim_Z K = +\infty$ and $\lim_Z \frac{\ln(K)}{\ln(n)} = 0$, implying that the $\lim_Z$ of the upper bound in Eq. (A78) is also $\frac{1}{1+Z}$, with a relative difference of order $1/[\ln(n)]^{1-\alpha}$ for finite n (the closer $\alpha > 0$ to 0, the more stringent this bound will be).

Finally, since $\lim_Z L_2 = 0$, we have shown that for any real $Z \geq 0$, $\lim_Z L_1^{\pm} = \lim_Z L_0(x) = \frac{1}{1+Z}$, which proves Theorem 2. Note that the two bounds obtained suggest that the relative error between $L_0(x)$ and $\frac{1}{1+Z}$ for finite large n and large $\|x - x_0\|^{-d}$ with $z(n, x_0)$ remaining close to Z is of order $1/\ln(n)$, or equivalently, of order $1/\ln(\|x - x_0\|)$.

Numerical simulations for the Lagrange function at finite n

In Fig. 1**c**, we illustrate numerically the scaling result of Theorem 2.

Note that, exploiting Theorem 2, we can use a simple heuristic argument to estimate the tail of the distribution of the random variable $W = w_0(x)$. Indeed, approximating $L_0(x)$ for finite but large n by its asymptotic form $\frac{1}{1+z_x(n,x_0)}$, with $z_x(n, x_0) = V_d \rho(x) n \log(n) \|x - x_0\|^d$, we obtain

$$\int_W^1 P(W')\, dW' \sim \int \rho(x_0)\, \theta \left(\frac{1}{1 + V_d \rho(x) n \log(n) \|x - x_0\|^d} - W \right) d^d x_0, \quad \text{(A79)}$$

$$\sim V_d \rho(x) \int_0^{+\infty} \theta \left(\frac{1}{1 + V_d \rho(x) n \log(n)\, u} - W \right) du, \quad \text{(A80)}$$

$$\sim \frac{1}{n \ln(n) W} \qquad \Longrightarrow P(W) \sim \frac{1}{n \ln(n) W^2}, \quad \text{(A81)}$$

where $\theta(.)$ is the Heaviside function. This heuristic result is again perfectly consistent with our guess (see the discussion below Theorem 1) that $P(W) = \frac{1}{W_n} p\left(\frac{W}{W_n}\right)$, with the scaling function p having the universal tail, $p(w) \underset{w \to +\infty}{\sim} w^{-2}$, and a scale $W_n \sim \frac{1}{n \ln(n)}$. Indeed, in this case and in the limit $n \to +\infty$, we obtain that $P(W) \sim \frac{1}{W_n} \left(\frac{W_n}{W}\right)^2 \sim \frac{W_n}{W^2} \sim \frac{1}{n \ln(n) W^2}$, which is identical to the result of Eq. (A81).

A.4 The variance term

We define the variance term $\mathcal{V}(x)$ as

$$\mathcal{V}(x) = \mathbb{E}\Big[\sum_{i=0}^n w_i^2(x)[y_i - f(x_i)]^2\Big] = \mathbb{E}_X\Big[\sum_{i=0}^n w_i^2(x)\sigma^2(x_i)\Big] = (n+1)\,\mathbb{E}\left[w_0^2(x)\sigma^2(x_0)\right]. \quad \text{(A82)}$$

If we first assume that $\sigma^2(x)$ is bounded by σ_0^2, we can readily bound $\mathcal{V}(x)$ using Theorem 1 with $\beta = 2$:

$$\mathcal{V}(x) \le (n+1)\sigma_0^2\, \mathbb{E}\left[w_0^2(x)\right]. \quad \text{(A83)}$$

Hence, for any $\varepsilon > 0$, there exists a constant $N_{x,\varepsilon}$, such that for $n \ge N_{x,\varepsilon}$, we obtain Theorem 3

$$\mathcal{V}(x) \le (1+\varepsilon)\frac{\sigma_0^2}{\ln(n)}. \quad \text{(A84)}$$

However, one can obtain an exact asymptotic equivalent for $\mathcal{V}(x)$ by assuming that σ^2 is continuous at x (with $\sigma^2(x) > 0$), while relaxing the boundedness condition. Indeed, we now assume the growth condition C^σ_{Growth}

$$\int \rho(y) \frac{\sigma^2(y)}{1 + \|y\|^{2d}}\, d^d y < \infty. \quad \text{(A85)}$$

Note that this condition can be satisfied even in the case where the mean variance $\int \rho(y)\sigma^2(y)\, d^d y$ is infinite.

Proceeding along the very same line as the proof of Theorem 1 in section A.2, we can write

$$\mathbb{E}\left[w_0^2(x)\sigma^2(x_0)\right] = \int_0^{+\infty} t\psi^n(x,t)\phi(x,t)\,dt, \tag{A86}$$

with

$$\phi(x,t) := \int \rho(x+y)\sigma^2(x+y)\frac{\mathrm{e}^{-\frac{t}{||y||^d}}}{||y||^{2d}}\,d^dy, \tag{A87}$$

which as a similar form as Eq. (A5), with $\beta = 2$. The condition of Eq. (A85) ensures that the integral defining $\phi(x,t)$ converges for all $t > 0$.

The continuity of σ^2 at x (and hence of $\rho\sigma^2$) and the fact the $\rho(x)\sigma^2(x) > 0$ implies the existence of a small enough $\lambda > 0$ such that the ball $B(x,\lambda) \subset \Omega^\circ$ and $||y|| \leq \lambda \implies |\rho(x+y)\sigma^2(x+y) - \rho(x)\sigma^2(x)| \leq \varepsilon\rho(x)\sigma^2(x)$, a property exploited for ρ in the proof of Theorem 1 (see Eq. (A18) and the paragraph above it), and which can now be used to efficiently bound $\phi(x,t)$. In addition, using the method of proof of Theorem 1 (see Eq. (A30)) also requires that $\int_{||y||\geq\lambda}\rho(y)\frac{\sigma^2(y)}{\|y\|^{2d}}\,d^dy < \infty$, which is ensured by the condition $C^{\sigma}_{\text{Growth}}$ of Eq. (A85). Apart from these details, one can proceed strictly along the proof and Theorem 1, leading to the proof of Theorem 4:

$$\mathcal{V}(x) \underset{n\to+\infty}{\sim} \frac{\sigma^2(x)}{\ln(n)}. \tag{A88}$$

Note that if $\sigma^2(x) = 0$, one can straightforwardly show that for any $\varepsilon > 0$, and for n large enough, one has

$$\mathcal{V}(x) \leq \frac{\varepsilon}{\ln(n)}, \tag{A89}$$

while a more optimal estimate can be easily obtained if one specifies how σ^2 vanishes at x.

A.5 The bias term

This section aims at proving Theorem 5, 6, and 7.

Assumptions

We first impose the following growth condition C^f_{Growth} for $f(x) := \mathbb{E}[Y \mid X = x]$:

$$\int \rho(y)\frac{f^2(y)}{(1+||y||^d)^2}\,d^dy < \infty, \tag{A90}$$

which is obviously satisfied if f is bounded. Since ρ is assumed to have a second moment, the condition C^f_{Growth} is also satisfied for any function satisfying $|f(x)| \leq A_f||y||^{d+1}$ for all y, such that $||y|| \geq R_f$, for some $R_f > 0$. Using the Cauchy-Schwartz inequality, we find that the condition C^f_{Growth} also implies that

$$\int \rho(y)\frac{|f(y)|}{1+||y||^d}\,d^dy < \infty. \tag{A91}$$

In addition, for any $x \in \Omega^\circ$ (so that $\rho(x) > 0$), we assume that there exists a neighborhood of x such that f satisfies a local Hölder condition. In other words, there exist $\delta_x > 0$, $K_x > 0$, and $\alpha_x > 0$, such that the ball $B(0, \delta_x) \subset \Omega$, and

$$||y|| \leq \delta_x \implies |f(x+y) - f(x)| \leq K_x ||y||^{\alpha_x}, \tag{A92}$$

which defines condition C^f_{Holder}.

Definition of the bias term and preparatory results

We define the bias term $\mathcal{B}(x)$ as

$$\mathcal{B}(x) = \mathbb{E}_X\Big[\Big(\sum_{i=0}^{n} w_i(x)[f(x_i) - f(x)]\Big)^2\Big] = (n+1)\mathcal{B}_1(x) + n(n+1)\mathcal{B}_2(x), \tag{A93}$$

$$\mathcal{B}_1(x) = \frac{1}{n+1}\mathbb{E}_X\Big[\sum_{i=0}^{n} w_i^2(x)[f(x_i) - f(x)]^2\Big], \tag{A94}$$

$$= \mathbb{E}_X\Big[w_0^2(x)[f(x_0) - f(x)]^2\Big], \tag{A95}$$

$$\mathcal{B}_2(x) = \frac{1}{n(n+1)}\mathbb{E}_X\Big[\sum_{0 \leq i < j \leq n} w_i(x)w_j(x)[f(x_i) - f(x)][f(x_j) - f(x)]\Big], \tag{A96}$$

$$= \mathbb{E}_X\Big[w_0(x)w_1(x)[f(x_0) - f(x)][f(x_1) - f(x)]\Big]. \tag{A97}$$

Exploiting again Eq. (A1) for $\beta = 2$ like we did in section A.2, we obtain

$$\mathcal{B}_1(x) = \int_0^{+\infty} t\, \psi^n(x,t)\chi_1(x,t)\, dt, \tag{A98}$$

where $\psi(x,t)$ is again the function defined in Eq. (A2), and where

$$\chi_1(x,t) := \int \rho(x+y) \mathrm{e}^{-\frac{t}{||y||^d}} \frac{(f(x+y) - f(x))^2}{||y||^{2d}}\, d^d y. \tag{A99}$$

For any $t > 0$, and under condition C^f_{Growth}, the integral defining $\chi_1(x,t)$ is well-defined. Moreover, $\chi_1(x,t)$ is a strictly positive and strictly decreasing function of $t > 0$.

Now, defining $u_i = ||x - x_i||^{-d}$, $i = 0, ..., n$ and exploiting again Eq. (A1) for $\beta = 2$, we can write

$$w_0(x)w_1(x) = u_0 u_1 \int_0^{\infty} t\, \mathrm{e}^{-(u_0+u_1)t - (\sum_{i=2}^{n} u_i)t}\, dt \tag{A100}$$

Now taking the expectation value over the $n+1$ independent variables, we obtain

$$\mathcal{B}_2(x) = \int_0^{+\infty} t\, \psi^{n-1}(x,t)\chi_2^2(x,t)\, dt, \tag{A101}$$

where

$$\chi_2(x,t) \coloneqq \int \rho(x+y)\mathrm{e}^{-\frac{t}{||y||^d}}\,\frac{f(x+y)-f(x)}{||y||^d}\,d^dy. \tag{A102}$$

Again, for any $t>0$, and under condition C^f_{Growth}, the integral defining $\chi_2(x,t)$ is well-defined. Note that, the integral defining $\chi_2(x,0)$ is well-behaved at $y=0$ under condition C^f_{Holder}. Indeed, for $||y||\le\delta_x$, we have $\frac{|f(x+y)-f(x)|}{||y||^d}\le K_x||y||^{-d+\alpha_x}$, which is integrable at $y=0$ in dimension d. Note that, if $f(x+y)-f(x)$ were only decaying as $const./\ln(||y||)$, then $|\chi_2(x,t)|\sim const.\ln(|\ln(t)|)\to+\infty$, when $t\to0$, and $\chi_2(x,0)$ would not exist (see the end of this section where we relax the local Hölder condition).

From now, we denote

$$\kappa(x) \coloneqq \chi_2(x,0) = \int \rho(x+y)\,\frac{f(x+y)-f(x)}{||y||^d}\,d^dy. \tag{A103}$$

Also note that $\kappa(x)=0$ is possible, even if f is not constant. For instance, if Ω is a sphere centered at x or $\Omega=\mathbb{R}^d$, if $\rho(x+y)=\hat\rho(||y||)$ is isotropic around x and, if $f_x: y\mapsto f(x+y)$ is an odd function of y, then we indeed have $\kappa(x)=0$ at the symmetry point x.

Upper bound for $\mathcal{B}_1(x)$

For $\varepsilon>0$, we define λ like in section A.2 and define $\eta=\min(\lambda,\delta_x)$, so that

$$\chi_1(x,t) \;\le\; (1+\varepsilon)K_x\rho(x)\int_{||y||\le\eta}\mathrm{e}^{-\frac{t}{||y||^d}}\,||y||^{2(\alpha_x-d)}\,d^dy+\Lambda_x, \tag{A104}$$

$$\Lambda_x \;=\; \int_{||y||\ge\eta}\rho(x+y)\,\frac{(f(x+y)-f(x))^2}{||y||^{2d}}\,d^dy, \tag{A105}$$

where the constant $\Lambda_x<\infty$ under condition C^f_{Growth}. The integral in Eq. (A104), can be written as

$$\int_{||y||\le\eta}\mathrm{e}^{-\frac{t}{||y||^d}}\,||y||^{2(\alpha_x-d)}\,d^dy \;=\; S_d\int_0^{\eta}\mathrm{e}^{-\frac{t}{r^d}}\,r^{2\alpha_x-d-1}\,dr, \tag{A106}$$

$$= V_d t^{\frac{2\alpha_x}{d}-1}\int_{\frac{t}{\eta^d}}^{+\infty}u^{-\frac{2\alpha_x}{d}}\,\mathrm{e}^{-u}\,du, \tag{A107}$$

Hence, we find that $\chi_1(x,t)$ is bounded for $\alpha_x>d/2$. For $\alpha_x<d/2$, and for $t<t_1$ small enough, there exists a constant $M(2\alpha_x/d)$ so that $\chi_1(x,t)\le M(2\alpha_x/d)t^{\frac{2\alpha_x}{d}-1}$. Finally, in the marginal case $\alpha_x=d/2$ and for $t<t_1$, we have $\chi_1(x,t)\le M(1)\ln(1/t)$, for some constant $M(1)$.

Now, exploiting again the upper bound of $\psi(x,t)$ obtained in section A.2 and repeating the steps to bound the integrals involving $\psi^n(x,t)$, we find that, for $\alpha_x\ne d/2$, $\mathcal{B}_1(x)$ is bounded up to a multiplicative constant by

$$\int_0^{t_1} t^{\min\left(1,\frac{2\alpha_x}{d}\right)}\,\mathrm{e}^{-nV_d\rho(x)t\ln\left(\frac{D_-}{t}\right)}\,dt \underset{n\to+\infty}{\sim} M'(2\alpha_x/d)\left(V_d\rho(x)n\ln(n)\right)^{-\min\left(2,\frac{2\alpha_x}{d}+1\right)}, \tag{A108}$$

where $M'(2\alpha_x/d)$ is a constant depending only on $2\alpha_x/d$. In the marginal case, $\alpha_x = d/2$, $\mathcal{B}_1(x)$ is bounded up to a multiplicative constant by $n^{-2}\ln(n)$.

In summary, we find that

$$(n+1)\mathcal{B}_1(x) = \begin{cases} O\left(n^{-\frac{2\alpha_x}{d}}(\ln(n))^{-1-\frac{2\alpha_x}{d}}\right), & \text{for } d > 2\alpha_x \\ O\left(n^{-1}(\ln(n))^{-1}\right), & \text{for } d = 2\alpha_x \\ O\left(n^{-1}(\ln(n))^{-2}\right), & \text{for } d < 2\alpha_x \end{cases} \tag{A109}$$

Asymptotic equivalent for $\mathcal{B}_2(x)$

Let us first assume that $\kappa(x) = \chi_2(x,0) \neq 0$. Then again, as shown in detail in section A.2, the integral defining $\mathcal{B}_2(x)$ is dominated by the small t region, and will be asymptotically equivalent to

$$\mathcal{B}_2(x) = \int_0^{+\infty} t\,\psi^{n-1}(x,t)\chi_2^2(x,t)\,dt, \tag{A110}$$

$$\underset{n\to+\infty}{\sim} \kappa^2(x)\int_0^{t_1} t\,\mathrm{e}^{-nV_d\rho(x)t\ln\left(\frac{D_\pm}{t}\right)}\,dt, \tag{A111}$$

$$\underset{n\to+\infty}{\sim} \left(\frac{\kappa(x)}{V_d\rho(x)n\ln(n)}\right)^2. \tag{A112}$$

On the other hand, if $\kappa(x) = 0$, one can bound $\chi_2(x,t)$ (up to a multiplicative constant) for $t \le t_1$ by the integral

$$\int_{||y||\le\eta}\left(1-\mathrm{e}^{-\frac{t}{||y||^d}}\right)||y||^{\alpha_x-d}\,d^dy = S_d\int_0^{\eta}\left(1-\mathrm{e}^{-\frac{t}{r^d}}\right)r^{\alpha_x-d}\,r^{d-1}\,dr, \tag{A113}$$

$$= V_d t^{\frac{\alpha_x}{d}}\int_{\frac{t}{\eta^d}}^{+\infty} u^{-1-\frac{\alpha_x}{d}}\left(1-\mathrm{e}^{-u}\right)du. \tag{A114}$$

Hence, for $\kappa(x) = 0$, we find that

$$n(n+1)\mathcal{B}_2(x) = O\left(n^{-\frac{2\alpha_x}{d}}(\ln(n))^{-2-\frac{2\alpha_x}{d}}\right). \tag{A115}$$

Asymptotic equivalent for the bias term $\mathcal{B}(x)$

In the generic case $\kappa(x) \neq 0$, we find that $(n+1)\mathcal{B}_1(x)$ is always dominated by $n(n+1)\mathcal{B}_2(x)$, and we find the following asymptotic equivalent for $\mathcal{B}(x) = (n+1)\mathcal{B}_1(x) + n(n+1)\mathcal{B}_2(x)$:

$$\mathcal{B}(x) \underset{n\to+\infty}{\sim} \left(\frac{\kappa(x)}{V_d\rho(x)\ln(n)}\right)^2. \tag{A116}$$

In the non-generic case $\kappa(x)=0$, the bound for $(n+1)\mathcal{B}_1(x)$ in Eq. (A109) is always more stringent than the bound for $n(n+1)\mathcal{B}_2(x)$ in Eq. (A115), leading to

$$\mathcal{B}(x)=\begin{cases} O\left(n^{-\frac{2\alpha_x}{d}}(\ln(n))^{-1-\frac{2\alpha_x}{d}}\right), & \text{for } d>2\alpha_x \\ O\left(n^{-1}(\ln(n))^{-1}\right), & \text{for } d=2\alpha_x \\ O\left(n^{-1}(\ln(n))^{-2}\right), & \text{for } d<2\alpha_x \end{cases}, \tag{A117}$$

which proves the statements made in Theorem 5.

Interpretation of the bias term $\mathcal{B}(x)$ for $\kappa(x)\neq 0$

Here, we assume the generic case $\kappa(x)\neq 0$ and define $\bar{f}(x)=\mathbb{E}\left[\hat{f}(x)\right]$. We have

$$\Delta(x) := \mathbb{E}\left[\sum_{i=0}^{n} w_i(x)(f(x_i)-f(x))\right] = \bar{f}(x)-f(x), \tag{A118}$$

$$\bar{f}(x) = \mathbb{E}\left[\sum_{i=0}^{n} w_i(x)f(x_i)\right] = (n+1)\mathbb{E}\left[w_0(x)f(x_0)\right]. \tag{A119}$$

By using another time Eq. (A1), we find that,

$$\Delta(x) = (n+1)\int_0^{+\infty}\psi^n(x,t)\chi_2(x,t)\,dt, \tag{A120}$$

$$\underset{n\to+\infty}{\sim} n\,\kappa(x)\int_0^{t_1} \mathrm{e}^{-nV_d\rho(x)t\ln\left(\frac{D_\pm}{t}\right)}\,dt, \tag{A121}$$

$$\underset{n\to+\infty}{\sim} \frac{\kappa(x)}{V_d\rho(x)\ln(n)}. \tag{A122}$$

Comparing this result to the one of Eq. (A116), we find that the bias $\mathcal{B}(x)$ is asymptotically dominated by the square of the difference $\Delta^2(x)$ between $\bar{f}(x)=\mathbb{E}\left[\hat{f}(x)\right]$ and $f(x)$:

$$\mathcal{B}(x) \underset{n\to+\infty}{\sim} \left(\mathbb{E}\left[\hat{f}(x)\right]-f(x)\right)^2, \tag{A123}$$

a statement made in Theorem 5.

Relaxing the local Hölder condition

We now only assume the condition $C^f_{\text{Cont.}}$ that f is continuous at x (but still assuming the growth conditions). We can now define δ_x such that the ball $B(x,\delta)\subset\Omega^\circ$ and $||y||\le\delta_x \implies |f(x+y)-f(x)|\le\varepsilon$. Then, the proof proceeds as above, but by replacing K_x by ε, α_x by 0, and by updating the bounds for $\chi_1(x,t)$ (for which this replacement is safe) and $\chi_2(x,t)$ (for which it is not). We now find that for $0<t\le t_1$,

with t_1 small enough

$$0 \leq \chi_1(x,t) \leq \varepsilon(1+2\varepsilon)V_d\rho(x)t^{-1}, \tag{A124}$$

$$|\chi_2(x,t)| \leq \varepsilon(1+2\varepsilon)V_d\rho(x)\ln\left(\frac{1}{t}\right). \tag{A125}$$

As already mentioned below Eq. (A102), where we provided an explicit counterexample, we see that relaxing the local Hölder condition does not guarantee anymore that $\lim_{t\to 0}|\chi_2(x,0)| < \infty$. With these new bounds, and carrying the rest of the calculation as in the previous sections, we ultimately find the following weaker result compared to Eq. (A116) and Eq. (A117):

$$\mathcal{B}(x) = o\left(\frac{1}{\ln(n)}\right), \tag{A126}$$

or equivalently, that for any $\varepsilon > 0$, there exists a constant $N_{x,\varepsilon}$ such that, for $n \geq N_{x,\varepsilon}$, we have

$$\mathcal{B}(x) \leq \frac{\varepsilon}{\ln(n)}, \tag{A127}$$

which proves Theorem 6.

The bias term at a point where $\rho(x) = 0$

This section aims at proving Theorem 7 expressing the lack of convergence of the estimator $\hat{f}(x)$ to $f(x)$, when $\rho(x) = 0$, and under mild conditions. Let us now consider a point $x \in \partial\Omega$ for which $\rho(x) = 0$, let us assume that there exists constants η_x, $\gamma_x > 0$, and $G_x > 0$, such that ρ satisfies the local Hölder condition at x

$$||y|| \leq \eta_x \implies \rho(x+y) \leq G_x||y||^{\gamma_x}. \tag{A128}$$

We will also assume that the growth condition of Eq. (A91) is satisfied. With these two conditions, $\kappa(x)$ defined in Eq. (A103) exists. The vanishing of ρ at x strongly affects the behavior of $\psi(x,t)$ in the limit $t \to 0$, which is not singular anymore:

$$1 - \psi(x,t) \underset{t\to 0}{\sim} t\int \rho(y)\|x-y\|^{-d}\, d^d y, \tag{A129}$$

where the convergence of the integral $\lambda(x) := \int \rho(y)\|x-y\|^{-d}\, d^d y$ is ensured by the local Hölder condition of ρ at x.

Let us now evaluate $\bar{f}(x) = \lim_{n\to+\infty}\mathbb{E}[\hat{f}(x)]$, the expectation value of the estimator $\hat{f}(x)$ in the limit $n \to +\infty$, introduced in Eq. (A119). First assuming, $\kappa(x) = \chi_2(x,0) \neq 0$, we obtain

$$\bar{f}(x) - f(x) = \lim_{n\to+\infty}(n+1)\int_0^{+\infty}\psi^n(x,t)\chi_2(x,t)\,dt, \tag{A130}$$

$$= \lim_{n\to+\infty} n\,\chi_2(x,0)\int_0^{t_1} \mathrm{e}^{n\,t\,\partial_t\psi(x,0)}\,dt, \tag{A131}$$

$$= \frac{\kappa(x)}{\lambda(x)}, \tag{A132}$$

which shows that the bias term does not vanish in the limit $n \to +\infty$. Eq. (A132) can be straightforwardly shown to remain valid when $\kappa(x) = 0$. Indeed, for any $\varepsilon > 0$ chosen arbitrarily small, we can choose t_1 small enough such that $|\chi_2(x,t)| \leq \varepsilon$ for $0 \leq t \leq t_1$, which leads to $|\bar{f}(x) - f(x)| \leq \varepsilon/\lambda(x)$.

Note that relaxing the local Hölder condition for ρ at x and only assuming the continuity of f at x and $\kappa(x) \neq 0$ is not enough to guarantee that $\bar{f}(x) \neq f(x)$. For instance, if $\rho(x+y) \sim_{y\to 0} \rho_0/\ln(1/||y||)$, and there exists a local solid angle $\omega_x > 0$ at x, one can show that $1 - \psi(x,t) \sim_{t\to 0} \omega_x S_d \rho_0\, t \ln(\ln(1/t))$, and the bias would still vanish in the limit $n \to +\infty$, with $\hat{f}(x) - f(x) \sim_{n\to+\infty} \kappa(x)/[\omega_x S_d \rho_0 \ln(\ln(n))]$.

A.6 Asymptotic equivalent for the regression risk

This section aims at proving Theorem 8. Under conditions $C^{\sigma}_{\mathrm{Growth}}$, C^{f}_{Growth}, and $C^{f}_{\mathrm{Cont.}}$, the results of Eq. (A88) and Eq. (A126) show that for $\rho(x)\sigma^2(x) > 0$ and ρ and σ^2 continuous at x, the bias term $\mathcal{B}(x)$ is always dominated by the variance term $\mathcal{V}(x)$ in the limit $n \to +\infty$. Thus, the excess regression risk satisfies

$$\mathbb{E}[(\hat{f}(x) - f(x))^2] \underset{n\to+\infty}{\sim} \frac{\sigma^2(x)}{\ln(n)}. \tag{A133}$$

As a consequence, the Hilbert kernel estimate converges pointwise to the regression function in probability. Indeed, for $\delta > 0$, there exists a constant $N_{x,\delta}$, such that

$$\mathbb{E}[(\hat{f}(x) - f(x))^2] \leq (1+\delta)\frac{\sigma^2(x)}{\ln(n)}, \tag{A134}$$

for $n \geq N_{x,\delta}$. Moreover, for any $\varepsilon > 0$, since $\mathbb{E}[(\hat{f}(x) - f(x))^2] \geq \varepsilon^2\, \mathbb{P}[|\hat{f}(x) - f(x)| \geq \varepsilon]$, we deduce the following Chebyshev bound, valid for $n \geq N_{x,\delta}$

$$\mathbb{P}[|\hat{f}(x) - f(x)| \geq \varepsilon] \leq \frac{1+\delta}{\varepsilon^2}\frac{\sigma^2(x)}{\ln(n)}. \tag{A135}$$

A.7 Rates for the plugin classifier

In the case of binary classification $Y \in \{0,1\}$ and $f(x) = \mathbb{P}[Y = 1 \mid X = x]$. Let $F\colon \mathbb{R}^d \to \{0,1\}$ denote the Bayes optimal classifier, defined by $F(x) := \theta(f(x) - 1/2)$ where $\theta(\cdot)$ is the Heaviside theta function. This classifier minimizes the risk $\mathcal{R}_{0/1}(h) := \mathbb{E}[\mathbb{1}_{\{h(X)\neq Y\}}] = \mathbb{P}[h(X) \neq Y]$ under zero-one loss. Given the regression estimator $\hat{f}$, we consider the plugin classifier $\hat{F}(x) = \theta(\hat{f}(x) - \frac{1}{2})$, and we will exploit the fact that

$$0 \leq \mathbb{E}[\mathcal{R}_{0/1}(\hat{F}(x))] - \mathcal{R}_{0/1}(F(x)) \leq 2\,\mathbb{E}[|\hat{f}(x) - f(x)|] \leq 2\sqrt{\mathbb{E}[(\hat{f}(x) - f(x))^2]} \tag{A136}$$

Proof of Eq. (A136)

For the sake of completeness, let us briefly prove the result of Eq. (A136). The rightmost inequality is simply obtained from the Cauchy-Schwartz inequality, and we hence focus on proving the first inequality. Obviously, Eq. (A136) is satisfied for $f(x) = 1/2$, for which $\mathbb{E}[\mathcal{R}_{0/1}(\hat{F}(x))] = \mathcal{R}_{0/1}(F(x)) = 1/2$.

If $f(x) > 1/2$, we have $F(x) = 1$, $\mathcal{R}_{0/1}(F(x)) = 1 - f(x)$, and

$$\begin{aligned} \mathbb{E}[\mathcal{R}_{0/1}(\hat{F}(x))] &= f(x)\mathbb{P}[\hat{f}(x) \le 1/2] + (1 - f(x))\mathbb{P}[\hat{f}(x) \ge 1/2], && \text{(A137)} \\ &= \mathcal{R}_{0/1}(F(x)) + (2f(x) - 1)\mathbb{P}[\hat{f}(x) \le 1/2], && \text{(A138)} \end{aligned}$$

which implies $\mathbb{E}[\mathcal{R}_{0/1}(\hat{F}(x))] \ge \mathcal{R}_{0/1}(F(x))$. Since $\mathbb{P}[\hat{f}(x) \le 1/2] = \mathbb{E}[\theta(1/2 - \hat{f}(x))]$, and using $\theta(1/2 - \hat{f}(x)) \le \frac{|\hat{f}(x)-f(x)|}{f(x)-1/2}$, valid for any $1/2 < f(x) \le 1$, we readily obtain Eq. (A136).

Similarly, in the case $f(x) < 1/2$, we have $F(x) = 0$, $\mathcal{R}_{0/1}(F(x)) = f(x)$, and

$$\mathbb{E}[\mathcal{R}_{0/1}(\hat{F}(x))] = \mathcal{R}_{0/1}(F(x)) + (1 - 2f(x))\mathbb{P}[\hat{f}(x) \ge 1/2]. \qquad \text{(A139)}$$

Since $\mathbb{P}[\hat{f}(x) \ge 1/2] = \mathbb{E}[\theta(\hat{f}(x) - 1/2)]$, and using $\theta(\hat{f}(x) - 1/2) \le \frac{|\hat{f}(x)-f(x)|}{1/2-f(x)}$, valid for any $0 \le f(x) < 1/2$, we again obtain Eq. (A136) in this case.

In fact, for any $\alpha > 0$, the inequalities $\theta(1/2 - \hat{f}(x)) \le \left(\frac{|\hat{f}(x)-f(x)|}{f(x)-1/2}\right)^{\alpha}$ and $\theta(\hat{f}(x) - 1/2) \le \left(\frac{|\hat{f}(x)-f(x)|}{1/2-f(x)}\right)^{\alpha}$ hold, respectively, for $f(x) > 1/2$ and $f(x) < 1/2$. Combining this remark with the use of the Hölder inequality leads to

$$\begin{aligned} \mathbb{E}[\mathcal{R}_{0/1}(\hat{F}(x))] - \mathcal{R}_{0/1}(F(x)) &\le 2|f(x) - 1/2|^{1-\alpha}\, \mathbb{E}\left[|\hat{f}(x) - f(x)|^{\alpha}\right], && \text{(A140)} \\ &\le 2|f(x) - 1/2|^{1-\alpha}\, \mathbb{E}\left[|\hat{f}(x) - f(x)|^{\frac{\alpha}{\beta}}\right]^{\beta}, && \text{(A141)} \end{aligned}$$

for any $0 < \beta \le 1$. In particular, for $0 < \alpha < 1$ and $\beta = \alpha/2$, we obtain

$$0 \le \mathbb{E}[\mathcal{R}_{0/1}(\hat{F}(x))] - \mathcal{R}_{0/1}(F(x)) \le 2|f(x) - 1/2|^{1-\alpha}\, \mathbb{E}\left[|\hat{f}(x) - f(x)|^2\right]^{\frac{\alpha}{2}}. \qquad \text{(A142)}$$

The interest of this last bound compared to the more classical bound of Eq. (A136) is to show explicitly the cancellation of the classification risk as $f(x) \to 1/2$, while still involving the regression risk $\mathbb{E}\left[|\hat{f}(x) - f(x)|^2\right]$ (to the power $\alpha/2 < 1/2$).

Bound for the classification risk

Now exploiting the results of section A.6 for the regression risk, and the two inequalities Eq. (A136) and Eq. (A142), we readily obtain Theorem 9.

A.8 Extrapolation behavior outside the support of ρ

This section aims at proving Theorem 10 characterizing the behavior of the regression estimator $\hat{f}$ outside the closed support $\bar{\Omega}$ of ρ (extrapolation).

Extrapolation estimator in the limit $n \to \infty$

We first assume the growth condition $\int \rho(y)\frac{|f(y)|}{1+\|y\|^d}\,d^dy < \infty$. For $x \in \mathbb{R}^d$ (i.e., not necessarily in Ω), we have quite generally

$$\mathbb{E}\left[\hat{f}(x)\right] = (n+1)\mathbb{E}\left[w_0(x)f(x)\right] = (n+1)\int_0^{+\infty} \psi^n(x,t)\chi(x,t)\,dt, \qquad \text{(A143)}$$

where $\psi(x,t)$ is again given by Eq. (A2) and

$$\chi(x,t) := \int \rho(x+y)f(x+y)\frac{\mathrm{e}^{-\frac{t}{||y||^d}}}{||y||^d}\,d^dy, \qquad \text{(A144)}$$

which is finite for any $t > 0$, thanks to the above growth condition for f.

Let us now assume that the point x is not in the closed support $\bar{\Omega}$ of the distribution ρ (which excludes the case $\Omega = \mathbb{R}^d$). Since the integral in Eq. (A143) is again dominated by its $t \to 0$ behavior, we have to evaluate $\psi(x,t)$ and $\chi(x,t)$ in this limit, like in the different proofs above. In fact, when $x \notin \bar{\Omega}$, the integral defining $\psi(x,t)$ and $\chi(x,t)$ are not singular anymore, and we obtain

$$1-\psi(x,t) \underset{t\to 0}{\sim} t\int \rho(y)\|x-y\|^{-d}\,d^dy, \qquad \text{(A145)}$$

$$\chi(x,0) = \int \rho(y)f(y)\|x-y\|^{-d}\,d^dy. \qquad \text{(A146)}$$

Note that $\psi(x,t)$ has the very same linear behavior as in Eq. (A129), when we assumed $x \in \partial\Omega$ with $\rho(x) = 0$, and a local Hölder condition for ρ at x.

Finally, by using the same method as in the previous sections to evaluate the integral of Eq. (A143) in the limit $n \to +\infty$, we obtain

$$\int_0^{+\infty} \psi^n(x,t)\chi(x,t)\,dt \underset{n\to+\infty}{\sim} \chi(x,0)\int_0^{t_1} \mathrm{e}^{n\,t\,\partial_t\psi(x,0)}\,dt, \qquad \text{(A147)}$$

$$\underset{n\to+\infty}{\sim} \frac{1}{n}\frac{\chi(x,0)}{|\partial_t\psi(x,0)|}, \qquad \text{(A148)}$$

which leads to the first result of Theorem 10:

$$\hat{f}_\infty(x) := \lim_{n\to+\infty} \mathbb{E}\left[\hat{f}(x)\right] = \frac{\int \rho(y)f(y)\|x-y\|^{-d}\,d^dy}{\int \rho(y)\|x-y\|^{-d}\,d^dy}. \qquad \text{(A149)}$$

Note that since the function $(x,y) \longmapsto \|x-y\|^{-d}$ is continuous at all points $x \notin \bar{\Omega}$, $y \in \Omega$, and thanks to the absolute convergence of the integrals defining $\hat{f}_\infty(x)$, standard methods show that $\hat{f}_\infty$ is continuous (in fact, infinitely differentiable) at all $x \notin \bar{\Omega}$.

Extrapolation far from Ω

Let us now investigate the behavior of $\hat{f}_\infty(x)$ when the distance $L := d(x,\Omega) = \inf\{||x-y||,\ y \in \Omega\} > 0$ between x and Ω goes to infinity, which can only happen for certain Ω, in particular, when Ω is bounded. We now assume the stronger condition, $\langle |f| \rangle := \int \rho(y)|f(y)|\, d^d y < \infty$, such that the ρ-mean of f, $\langle f \rangle := \int \rho(y) f(y)\, d^d y$, is finite. We consider a point $y_0 \in \Omega$, so that $||x - y_0|| \geq L > 0$, and we will exploit the following inequality, valid for any $y \in \Omega$ satisfying $||y - y_0|| \leq R$, with $R > 0$:

$$0 \leq 1 - \frac{L^d}{||x-y||^d} \leq \frac{||x-y||^d - L^d}{L^d} \leq \frac{(L+R)^d - L^d}{L^d} \leq \mathrm{e}^{\frac{dR}{L}} - 1. \tag{A150}$$

Now, for a given $\varepsilon > 0$, there exist $R > 0$ large enough such that $\int_{\|y-y_0\| \geq R} \rho(y)\, d^d y \leq \varepsilon/2$ and $\int_{\|y-y_0\| \geq R} \rho(y)|f(y)|\, d^d y \leq \varepsilon/2$. Then, for such a R, we consider L large enough such that the above bound satisfies $\mathrm{e}^{\frac{dR}{L}} - 1 \leq \varepsilon \min(1/\langle |f| \rangle, 1)/2$. We then obtain

$$\left| L^d \int \rho(y) f(y) \|x-y\|^{-d}\, d^d y - \langle f \rangle \right| \leq \left(\mathrm{e}^{\frac{dR}{L}} - 1 \right) \int_{||y-y_0|| \leq R} \rho(y)|f(y)|\, d^d y \tag{A151}$$

$$+ \int_{\|y-y_0\| \geq R} \rho(y)|f(y)|\, d^d y, \tag{A152}$$

$$\leq \frac{\varepsilon}{2\langle |f| \rangle} \times \langle |f| \rangle + \frac{\varepsilon}{2} \leq \varepsilon, \tag{A153}$$

which shows that under the condition $\langle |f| \rangle < \infty$, we have

$$\lim_{d(x,\Omega) \to +\infty} d^d(x, \Omega) \int \rho(y) f(y) \|x-y\|^{-d}\, d^d y = \langle f \rangle. \tag{A154}$$

Similarly, one can show that

$$\lim_{d(x,\Omega) \to +\infty} d^d(x, \Omega) \int \rho(y) \|x-y\|^{-d}\, d^d y = \int \rho(y)\, d^d y = 1. \tag{A155}$$

Finally, we obtain the second result of Theorem 10,

$$\lim_{d(x,\Omega) \to +\infty} \hat{f}_\infty(x) = \langle f \rangle. \tag{A156}$$

Continuity of the extrapolation

We now consider $x \notin \bar{\Omega}$ and $y_0 \in \partial\Omega$, but such that $\rho(y_0) > 0$ (i.e., $y_0 \in \partial\Omega \cap \Omega$), and we note $l := ||x - y_0|| > 0$. We assume the continuity at y_0 of ρ and f as seen as functions restricted to Ω, i.e., $\lim_{y \in \Omega \to y_0} \rho(y) = \rho(y_0)$ and $\lim_{y \in \Omega \to y_0} f(y) = f(y_0)$. Hence, for any $0 < \varepsilon < 1$, there exists $\delta > 0$ small enough such that $y \in \Omega$ and $||y - y_0|| \leq \delta \implies |\rho(y_0) - \rho(y)| \leq \varepsilon$ and $|\rho(y_0) f(y_0) - \rho(y) f(y)| \leq \varepsilon$. Since we intend to take $l > 0$ arbitrary small, we can impose $l < \delta/2$.

We will also assume that $\partial\Omega$ is smooth enough near y_0, such that there exists a strictly positive local solid angle ω_0 defined by

$$\omega_0 = \lim_{r\to 0} \frac{1}{V_d \rho(y_0) r^d} \int_{\|y-y_0\|\le r} \rho(y)\, d^d y = \lim_{r\to 0} \frac{1}{V_d r^d} \int_{y\in\Omega/\|y-y_0\|\le r} d^d y, \tag{A157}$$

where the second inequality results from the continuity of ρ at y_0 and the fact that $\rho(y_0) > 0$. If $y_0 \in \Omega^\circ$, we have $\omega_0 = 1$, while for $y_0 \in \partial\Omega$, we have generally $0 \le \omega_0 \le 1$. Although we will assume $\omega_0 > 0$ for our proof below, we note that $\omega_0 = 0$ or $\omega_0 = 1$ can happen for $y_0 \in \partial\Omega$. For instance, we can consider Ω_0, $\Omega_1 \subset \mathbb{R}^2$ respectively defined by $\Omega_0 = \{(x_1, x_2) \in \mathbb{R}^2/x_1 \ge 0,\ |x_2| \le x_1^2\}$ and $\Omega_1 = \{(x_1, x_2) \in \mathbb{R}^2/x_1 \le 0\} \cup \{(x_1, x_2) \in \mathbb{R}^2/x_1 \ge 0,\ |x_2| \ge x_1^2\}$. Then, it is clear that the local solid angle at the origin $O = (0, 0)$ is respectively $\omega_0 = 0$ and $\omega_0 = 1$. Also note that if x is on the surface of a sphere or on the interior of a face of a hypercube (and in general, when the boundary near x is locally a hyperplane; the generic case), we have $\omega_x = \frac{1}{2}$. If x is a corner of the hypercube, we have $\omega_x = \frac{1}{2^d}$.

Returning to our proof, and exploiting Eq. (A157), we consider δ small enough such that for all $0 \le r \le \delta$, we have

$$\left| \int_{y\in\Omega/\|y-y_0\|\le r} d^d y - \omega_0 V_d\, r^d \right| \le \varepsilon\, \omega_0 V_d\, r^d. \tag{A158}$$

We can now use these preliminaries to obtain

$$(\rho(y_0) f(y_0) - \varepsilon) J(x) - C \le \int \rho(y) f(y) \|x-y\|^{-d}\, d^d y \le (\rho(y_0) f(y_0) + \varepsilon) J(x) + C, \tag{A159}$$

$$(\rho(y_0) - \varepsilon) J(x) - C' \le \int \rho(y) \|x-y\|^{-d}\, d^d y \le (\rho(y_0) + \varepsilon) J(x) + C', \tag{A160}$$

with

$$J(x) := \int_{y\in\Omega\,/\,||y-y_0||\le\delta} \|x-y\|^{-d}\, d^d y, \tag{A161}$$

$$C = \left(\frac{2}{\delta}\right)^2 \int_{||y-y_0||\ge\delta} \rho(y) |f(y)|\, d^d y, \tag{A162}$$

$$C' = \left(\frac{2}{\delta}\right)^2. \tag{A163}$$

Let us now show that $\lim_{l\to 0} J(x) = +\infty$. We define $N := [\delta/l] \ge 2$, where $[\,.\,]$ is the integer part, and we have $N \ge 2$, since we have imposed $l < \delta/2$. For $n \in \mathbb{N} \ge 1$, we define,

$$I_n := \int_{y\in\Omega/||y-y_0||\le\delta/n} d^d y, \tag{A164}$$

and note that we have

$$I_n - I_{n+1} = \int_{\substack{y\in\Omega / ||y-y_0||\leq \delta/n, \\ ||y-y_0||\geq \delta/(n+1)}} d^d y, \tag{A165}$$

$$\left| I_n - \omega_0 V_d \left(\frac{\delta}{n}\right)^d \right| \leq \varepsilon\, \omega_0 V_d \left(\frac{\delta}{n}\right)^d . \tag{A166}$$

We can then write

$$J(x) \geq \sum_{n=1}^{N} \frac{1}{\left(l+\frac{\delta}{n}\right)^d} (I_n - I_{n+1}), \tag{A167}$$

$$\geq \sum_{n=1}^{N} \left(\frac{1}{\left(l+\frac{\delta}{n+1}\right)^d} - \frac{1}{\left(l+\frac{\delta}{n}\right)^d} \right) I_{n+1} + \frac{I_1}{(l+\delta)^d} - \frac{I_{N+1}}{\left(l+\frac{\delta}{N+1}\right)^d}. \tag{A168}$$

We have

$$\frac{I_1}{(l+\delta)^d} - \frac{I_{N+1}}{\left(l+\frac{\delta}{N+1}\right)^d} \geq \omega_0 V_d \left((1-\varepsilon)\frac{1}{\left(1+\frac{l}{\delta}\right)^d} - (1+\varepsilon)\frac{1}{\left(1+\frac{(N+1)l}{\delta}\right)^d} \right), \tag{A169}$$

$$\geq \omega_0 V_d \left((1-\varepsilon)\frac{2^d}{3^d} - (1+\varepsilon) \right) =: C'', \tag{A170}$$

which defines the constant C''. Now using Eq. (A166), $l < \delta/2$, $N = [\delta/l]$, and the fact that $(1+u)^d - 1 \geq d\,u$, for any $u \geq 0$, we obtain

$$J(x) \geq (1-\varepsilon)\, \omega_0 V_d \sum_{n=1}^{N} \frac{1}{\left(1+\frac{(n+1)l}{\delta}\right)^d} \left(\left(\frac{l+\frac{\delta}{n}}{l+\frac{\delta}{n+1}} \right)^d - 1 \right) + C'', \tag{A171}$$

$$\geq (1-\varepsilon)\, \omega_0 S_d \sum_{n=1}^{N} \frac{1}{\left(1+\frac{(n+1)l}{\delta}\right)^{d+1}} \frac{1}{n} + C'', \tag{A172}$$

$$\geq \frac{(1-\varepsilon)\, \omega_0\, S_d}{\left(1+\frac{(N+1)l}{\delta}\right)^{d+1}} \ln(N-1) + C'', \tag{A173}$$

$$\geq (1-\varepsilon)\, \omega_0 \left(\frac{2}{5}\right)^{d+1} S_d \ln\left(\frac{\delta}{l} - 2\right) + C''. \tag{A174}$$

We hence have shown that $\lim_{l\to 0} J(x) = +\infty$. Note that we can obtain an upper bound for $J(x)$ similar to Eq. (A172) in a similar way as above, and with a bit more work, it is straightforward to show that we in fact have $J(x) \sim_{l\to 0} \omega_0\, S_d \ln\left(\frac{\delta}{l}\right)$, a result that we will not need here.

Now, using Eq. (A159) and Eq. (A160) and the fact that $\lim_{l\to 0} J(x) = +\infty$, we find that

$$\int \rho(y) f(y) \|x-y\|^{-d}\, d^d y \underset{l\to 0}{\sim} \rho(y_0) f(y_0) J(x), \tag{A175}$$

$$\int \rho(y) \|x-y\|^{-d}\, d^d y \underset{l\to 0}{\sim} \rho(y_0) J(x), \tag{A176}$$

for $f(y_0) \neq 0$ (remember that $\rho(y_0) > 0$), while for $f(y_0) = 0$, we obtain $\int \rho(y) f(y) \|x-y\|^{-d}\, d^d y = o(J(x))$. Finally, we have shown that

$$\lim_{x\notin\bar{\Omega},\, x\to y_0} \hat{f}_\infty(x) = f(y_0), \tag{A177}$$

establishing the continuity of the extrapolation and the last part of Theorem 10.