$\mathcal{H}$ Matrix

For completeness, we discuss the hierarchical matrix technique. Especially we give a detailed description on the storage format, construction, fast matrix vector multiplication routine and LU decomposition. We later show how to construct the $\mathcal{H}$ matrix from kernels.

The discretization of the jump diffusion part $\int_{\mathbb{R}} (u(x+y) - u(y))dy$ will usually lead to a dense matrix, which typically requires $\mathcal{O}(N^2)$ storage and has $\mathcal{O}(N^2)$ complexity for matrix-vector multiplication, $\mathcal{O}(N^3)$ for LU decomposition. Many techniques, such as the panel clustering methods and the fast multipole methods were developed. Later $\mathcal{H}$-matrix was considered by W. Hackbusch and many variations of hierarchical matrices have been intensively studied by researchers. $\mathcal{H}$-matrices can reduces the storage and arithmetics to nearly optimal complexity $\mathcal{O}(N)$ up to $\log N$ scaling. It relies on the fact that the kernel functions are smooth in the off diagonal.

0.1 Construction and Storage

The construction of the $\mathcal{H}$ matrices can be best described in terms of matrix indices and the geometric points. Each entry A_{ij} represents the interaction between two nodes x_i and x_j. Let I, $J \subset \mathbb{N}$ be row and column index sets, then $A_{IJ} = (a_{ij})_{i \in I, j \in J}$ describes the interaction between a cluster $X_I = \{x_i\}_{i \in I}$ and another cluster $X_J = \{x_j\}_{j \in I}$. The interaction kernel function $k(x, y)$ is assumed to be smooth for sufficiently large $|x - y|$.

Typically, it requires $\mathcal{O}(|I||J|)$ complexity to store the interaction data. However, if we assume that $I \subset J = \emptyset$ and geometrically the clusters X_I, X_J are separate in the sense of admissibility

Definition 0.1. For two sets of indices I and J and the associated cluster X_I, X_J; assume that the kernel is asymptotically smooth, the admissibility condition is given by

$$\min\{\text{diam}(X_I), \text{diam}(X_J)\} \leq \eta \text{dist}(X_I, X_J) \tag{1}$$

where $\text{diam}(X_I) = \max_{x_i, x_j \in X} |x_i - x_j|$ AND $\text{dist}(x_i, x_j) = \min_{x_i \in X_I, x_j \in X_J} |x_i - x_j|$. If the condition eq. (1) is not satisfied, we say X_I and X_J or I and J are inadmissible.

In our numerical examples, we use $\eta = 1$, which indicates adjacent clusters are inadmissible since the distance is always zero.

The admissible blocks usually have low rank structures. This is best illustrated by an example. Suppose $k(x, y) = \frac{1}{|x-y|^2}$, and further assume $x \in X_I$, $y \in X_J$. Assume X_I and Y_J are inadmissible, and $\bar{x} \in \mathcal{X}$, where $\mathcal{X}$ is the convex hull of X_I. Then we have

$$\frac{1}{|x-y|^2} = \frac{1}{|x - \bar{x} - (y - \bar{x})|^2} = \frac{1}{|y - \bar{x}|^2 \left| \frac{x-\bar{x}}{y-\bar{x}} - 1 \right|^2}$$

Since

$$\left| \frac{x - \bar{x}}{y - \bar{x}} \right| < 1$$

we have

$$\frac{1}{|y - \bar{x}|^2 \left|\frac{x - \bar{x}}{y - \bar{x}} - 1\right|^2} = \frac{1}{|y - \bar{x}|^2} \left(1 + \frac{x - \bar{x}}{y - \bar{x}} + \left(\frac{x - \bar{x}}{y - \bar{x}}\right)^2 + \ldots\right)$$

which is in the form of

$$\frac{1}{|x - y|^2} = \sum_{n=0}^{\infty} \alpha_n(x - \bar{x})\beta_n(y - \bar{x}) \tag{2}$$

Then series is convergent and therefore the residual term will decay. It is possible to approximate $\frac{1}{|x-y|^2}$ with a few terms

$$\frac{1}{|x - y|^2} \approx \sum_{n=0}^{r} \alpha_n(x - \bar{x})\beta_n(y - \bar{x}) \tag{3}$$

And therefore the interaction matrix for the cluster A_{IJ} is

$$A_{IJ} = \left(\frac{1}{|x_i - x_j|^2}\right)_{i \in I, j \in J} = \left(\sum_{n=0}^{r} \alpha_n(x_i - \bar{x})\beta_n(x_j - \bar{x})\right)_{i \in I, j \in J} = UV' \tag{4}$$

where
$$U = \begin{bmatrix} \alpha_0(x_i - \bar{x}) & \alpha_1(x_i - \bar{x}) & \ldots & \alpha_r(x_i - \bar{x}) \end{bmatrix} \tag{5}$$
$$V = \begin{bmatrix} \beta_0(x_i - \bar{x}) & \beta_1(x_i - \bar{x}) & \ldots & \beta_r(x_i - \bar{x}) \end{bmatrix} \tag{6}$$

If $r \ll |I| \wedge |J|$, we have achieved matrix compression using a low rank representation.

The idea of the hierarchical matrix is then to classify each block A_{IJ} into three types

- Full matrix. In this case, A_{IJ} is represented using fully populated matrices.

- Low rank matrix. In the case I and J are admissible, we can store the block A_{IJ} in the form of low rank matrices. This will help us save storage and computational cost.

- $\mathcal{H}$-matrix. For the blocks that are neither low rank matrix nor small enough to become a full matrix, it is further divided into subblocks (for example, via quadtree structure).

The $\mathcal{H}$-matrix will be stored in a hierarchical format.

0.2 Matrix Vector Multiplication

One advantage of the $\mathcal{H}$ matrix is that the matrix vector multiplication is cheap. The matrix vector multiplication of $\mathcal{H}$-matrix can be described through the rule of the operator for three different kinds of subblocks

- Full matrix. In this case, the normal dense matrix vector multiplication is used.

2

- Low rank matrix. The operator can be carried out quite efficiently via

$$(UV')x = U(V'x) \tag{7}$$

note $V'x$ is a $r \times 1$ vector.

- $\mathcal{H}$-matrix. If the subblock is

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{8}$$

the matrix vector multiplication will be carried out recursively, i.e.

$$Bx = \begin{bmatrix} B_{11}x_1 + B_{12}x_2 \\ B_{21}x_1 + B_{22}x_2 \end{bmatrix} \tag{9}$$

0.3 LU Decomposition

$\mathcal{H}$-LU can be done in $\mathcal{H}$-matrix format and recursively in computational cost $\mathcal{O}(N)$ up to a $\log N$ scaling compared to dense LU in $\mathcal{O}(N^3)$.

We need to define a triangular solver which solves $AX = B$ for lower triangular matrix or $XA = B$ or upper triangular matrix. The matrices are either $\mathcal{H}$-matrix or full matrix. We only need to consider the lower triangular cases since in the latter case by transposition $A'X' = B'$, we reduce the problem to former.

The triangular solver will work differently for different situations.

- If B is a full matrix, then X is a full matrix and $X = A^{-1}B$. Here A is converted to a full matrix.

- If B is a low rank matrix, $B = B_1 B_2'$, then X is also a low rank matrix $X = (A^{-1}B_1)B_2'$.

- If A and B are both hierarchical matrices

$$\begin{bmatrix} A_{11} & \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

Then we will first solve $A_{11}X_{11} = B_{11}$ and $A_{11}X_{12} = B_{12}$. Then we solve

$$A_{22}X_{21} = B_{21} - A_{21}X_{11} \quad A_{22}X_{22} = B_{22} - A_{21}X_{12} \tag{10}$$

The LU decomposition also works differently for different types of matrices. Again only full matrices and $\mathcal{H}$ matrices are considered.

For full matrices, the standard dense LU is adopted. For $\mathcal{H}$ matrices,

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} L_{11} & \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ & U_{22} \end{bmatrix}$$

The algorithm will works as follows

- LU decomposition of $A_{11} = L_{11}U_{11}$

- Triangular solve $L_{11}U_{12} = A_{12}$ (lower triangular, U_{12} is the unknown)

- Triangular solve $L_{21}U_{11} = A_{21}$ (upper triangular, L_{21} is the unknown)

- LU decomposition of $A_{22} - L_{21}U_{12} = L_{22}U_{22}$

The LU decomposition can also be performed in an in-place way, which will save storage.

Lévy Process

Consider a given probability space $(\Omega, \mathcal{F}, P)$. A Lévy process $\{X_t\}_{t \geq 0}$ taking values in $\mathbb{R}^d$ is defined as a stochastic process with stationary and independent increments. In addition, we assume $X_0 = 0$ with probability 1.

By independent, we mean for any distinct time $0 \leq t_1 < t_2 < \ldots < t_n$, we have $X_{t_1}, X_{t_2 - t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are all independent.

By stationary, for any $0 \leq s < t < \infty$, the probability distribution of $X_t - X_s$ is the same as X_{t-s}.

One remarkable property of the Lévy process is that any Lévy process has a specific form of the characteristic function, called *Lévy-Khintchine formula*

$$\mathbb{E}(e^{\mathrm{i}(\xi, X_t)}) = e^{t\eta(\xi)} \tag{11}$$

where

$$\eta(\xi) = \mathrm{i}(b, \xi) - \frac{1}{2}(\xi, a\xi) + \int_{\mathbb{R}^d \setminus \{0\}} \left[e^{\mathrm{i}(\xi, y)} - 1 - \mathrm{i}(\xi, y)\mathbf{1}[0 < |y| < 1](y) \right] dy \tag{12}$$

here $b \in \mathbb{R}^d$, a is a positive definite symmetric matrix in $\mathbb{R}^{d \times d}$, and ν is a Lévy measure which satisfies

$$\int_{\mathbb{R}^d \setminus \{0\}} \min\{1, |y|^2\}\nu(dy) < \infty \tag{13}$$

In the case $\nu \equiv 0$, we obtain the Gaussian process. In the case $\int_{\mathbb{R}^d} \left[e^{\mathrm{i}(\xi, y)} - 1 \right] dy$ is well defined, we can omit the term $\mathrm{i}(\xi, y)\mathbf{1}[0 < |y| < 1](y)$.

In the case $\nu < \infty$, the Lévy process has the decomposition

$$X_t = bt + \sqrt{a}B_t + \sum_{0 \leq s \leq t} J_s \tag{14}$$

where J_s is the jump at time s. To be precise, define

$$N(t, A) = \#\{0 \leq s \leq t : J_s \in A\} \tag{15}$$

if t and A is fixed, $N(t, A)$ is a random variable; if t and $w \in \Omega$ is fixed, $N(t, \cdot)(w)$ is a measure; if A is fixed, $N(\cdot, A)$ is a Poisson process with intensity $\nu(A)$. Therefore, we can also write

$$\sum_{0 \leq s \leq t} J_s = \int_{\mathbb{R}^d - \{0\}} xN(t, dx) \tag{16}$$

To end this section, we provide a third view of the Lévy process. Consider the semigroup

$$(T_t f)(x) = \mathbb{E}(f(X_t + x)) \tag{17}$$

Then the infinitesimal generator will have the form

$$(Af)(x) = b^i(\partial_i f)(x) + \frac{1}{2}a^{ij}(\partial_i \partial_j f)(x) +$$
$$\int_{\mathbb{R}^d \setminus \{0\}} [f(x + y) - f(x) - y^i(\partial_i f)(x)\mathbf{1}_{0 < |y| < 1}(y)]\nu(dy) \tag{18}$$

Remark 1. Another definition of the infinitesimal generator is through the Fourier transform

$$(Af)(x) = \lim_{t \to 0+} \frac{P_t f - f}{t} \tag{19}$$

where $P_t f = f \star p_t$ and $\mathcal{F} p_t(\xi) = e^{-t\eta(\xi)}$.

To see this, consider the 1D case and without the adjustment term $-y^i(\partial_i f)(x)\mathbf{1}_{0<|y|<1}(y)$. By taking the Fourier transform of $(Af)(x)$, we have

$$\mathcal{F}(Af)(\xi) = \mathrm{i}\xi b \hat{f}(\xi) - \frac{1}{2}a\xi^2 \hat{f}(\xi) + \int_{-\infty}^{+\infty} (\hat{f}(\xi)e^{\mathrm{i}y\xi} - \hat{f}(\xi))\nu(y)dy$$

$$= \left(\mathrm{i}\xi b - \frac{1}{2}a\xi^2 + \int_{-\infty}^{+\infty} (e^{\mathrm{i}y\xi} - 1)\nu(y)dy \right) \hat{f}(\xi) \tag{20}$$

this is exactly the expression we see in eq. (12).

One the other hand,

$$\mathcal{F}\left[\lim_{t \to 0+} \frac{P_t f - f}{t} \right] = \lim_{t \to 0+} \frac{e^{\eta(\xi)t} \hat{f}(\xi) - \hat{f}(\xi)}{t} = \eta(\xi)$$

which coincides with eq. (20).

Again, we can drop the extra term $y^i(\partial_i f)(x)\mathbf{1}_{0<|y|<1}(y)$ if $\int_{\mathbb{R}^d}[f(x+y) - f(x)]\nu(dy)$ is well defined. We call the Lévy measure is *regular* if this condition is fulfilled.

Fokker Planck Equation

Let T_t be the semigroup associated with Lévy process, and the associated infinitesimal generator is

$$(Af)(x) = c(x)f(x) + b^i(\partial_i f)(x) + \frac{1}{2}a^{ij}(\partial_i \partial_j f)(x) +$$

$$\int_{\mathbb{R}^d \setminus \{0\}} [f(x+y) - f(x) - y^i(\partial_i f)(x)\mathbf{1}_{0<|y|<1}(y)]\nu(dy) \tag{21}$$

we consider the transition measures $p_t(x)$ associated with T_t. Here p_t is absolutely continuous with respect to Lebesgue measure. Define the adjoint operator A^* of A, which satisfies

$$\int_{\mathbb{R}^d} (Af)(y)p_t(y)dy = \int_{\mathbb{R}^d} f(y)A^* p_t(y)dy \tag{22}$$

for all $f \in C_c^\infty(\mathbb{R}^d)$.

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.750.9443&rep=rep1&type=pdf

In general, there is no nice form for A^*. However, in the case $c(x)$, $b(x)$ and $a(x)$ are all constant, we have

$$A^* p_t(x) = c(\partial_i p_t)(x) - b^i(\partial_i p_t)(x) + \frac{1}{2}a^{ij}\partial_i \partial_j p_t +$$

$$\int_{\mathbb{R}^d} \left[p_t(x-y) - p_t(x) + y^i(\partial_i p_t)(x)\mathbf{1}_{0<|y|<1}(y) \right] \nu(dy) \tag{23}$$

5

The Fokker-Planck equation, or Kolmogorov forward equation, is

$$\frac{\partial p_t(x)}{\partial t} = A^* p_t(x) \quad p_0(x) = \delta(x) \tag{24}$$

We will consider the forward or backward equation driven by the Lévy process, where the model problem in 1D can be stated as a convection-diffusion integro-differential equation

$$u_t = au_{xx} + bu_x + cu + \mathcal{L}u \quad x \in \mathbb{R}, t \in (0,1) \tag{25}$$

where $a \geq 0$, $c \leq 0$, and

$$\mathcal{L}u = \int_{\mathbb{R}} (u(x+y) - u(x) - u'(x)\mathbf{1}_{0<|y|<1}(y)y)\nu(y)dy \tag{26}$$

We consider the case where $\nu(y) < \infty$ and thus the term $u'(x)\mathbf{1}_{0<|y|<1}(y)y$ is not needed. The case $\nu(y)$ might grow to infinity at $y = 0$ will be discussed later. To compute numerically the integral term, we need to restrict the computational domain to a bounded interval Ω

$$\mathcal{L}u \approx \int_{B_l}^{B_r} (u(x+y) - u(x) - u'(x)\mathbf{1}_{0<|y|<1}(y)y)\nu(y)dy \tag{27}$$

In fact, it is proved in ??? that if $\nu(dy)$ is semi-heavy, i.e., there exists $\alpha_r, \alpha_l > 0$, such that $\int_1^\infty e^{(1+\alpha_r)y}\nu(dy) < \infty$, and $\int_{-\infty}^{-1} |y|e^{\alpha_l|y|}\nu(dy) < \infty$, the solution $\tilde{u}(x,t)$ obtained using the truncated integral will satisfy

$$|u(x,t) - \tilde{u}(x,t)| = \mathcal{O}(e^{-\alpha_l|B_l|} + e^{-\alpha_r|B_r|}) \tag{28}$$

Therefore, the discretization scheme for $\mathcal{L}$ on uniform grid will be

$$(\mathcal{L}u)(jh) \approx \sum_{j \in \mathcal{I}} u_{i+j}\nu_j h - u_i \lambda h \quad \lambda = \sum_{j \neq 0, j \in \mathcal{I}} \nu_j \tag{29}$$

Here $\nu_j = \nu(jh)$ and

$$\mathcal{I} = \{i : ih \in \Omega\} \tag{30}$$

Consider the Merton model with Gaussian jumps, i.e., the Lévy density can be represented as

$$\nu(x) = e^{-\varepsilon^2 x^2} \tag{31}$$

We consider the kernel function $k(x,y)$ associated with the density

$$k(x,y) = \nu(x-y) = e^{-\varepsilon^2(x-y)^2} \tag{32}$$

Assume that $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $\mathcal{X} \cap \mathcal{Y} = \emptyset$, and let $\bar{x} \in \mathcal{X}$. Denote $t_0 = x - \bar{x}$, and $t = y - \bar{x}$, then by assumption we have $|t| > |t_0|$. From Taylor expansion we have

$$e^{-\varepsilon^2(x-y)^2} = e^{-\varepsilon^2(t-t_0)^2} \tag{33}$$

$$= e^{-\varepsilon^2 t^2 \left(1-\frac{t_0}{t}\right)^2} = e^{-\varepsilon^2 t^2 - \varepsilon^2 t_0^2 + 2\varepsilon^2 t_0 t} \tag{34}$$

$$= e^{-\varepsilon^2 t^2 - \varepsilon^2 t_0^2} \left(1 + 2\varepsilon^2 t_0 t + \frac{(2\varepsilon^2 t_0 t)^2}{2} + \frac{(2\varepsilon^2 t_0 t)^3}{3!} + \dots\right) \tag{35}$$

6

Thus we have

$$\alpha_n(t) = \frac{2^n \varepsilon^{2n} e^{-\varepsilon^2 t^2} t^n}{n!} \quad \beta_n(t) = e^{-\varepsilon^2 t^2} t^n \tag{36}$$

we will have

$$e^{-\varepsilon^2 (x-y)^2} = \sum_{n=0}^{\infty} \alpha_n(t_0)\beta_n(t) \tag{37}$$

Lemma 1. Assume $\mathcal{X}$, $\mathcal{Y}$ are two disjoint set in Ω and $\mathrm{diam}(\Omega) = D$. Let $\delta > 0$ be any positive constant, then if

$$r > \max\left\{\log_2\left(\frac{e^{2\varepsilon^2 D^2}}{\delta}\right) - 1, 12\varepsilon^2 D^2 - 1\right\} \tag{38}$$

we have

$$\left| e^{-\varepsilon^2 (x-y)^2} - \sum_{n=0}^{r} \alpha_n(x - \bar{x})\beta_n(y - \bar{x}) \right| < \delta \tag{39}$$

for any $\bar{x} \in \mathcal{X}$.

Proof. Note that

$$\left| e^{-\varepsilon^2 (x-y)^2} - \sum_{n=0}^{r} \alpha_n(x - \bar{x})\beta_n(y - \bar{x}) \right| \tag{40}$$

$$= e^{-\varepsilon^2 t^2 - \varepsilon^2 t_0^2}\left(\frac{(2\varepsilon^2 t_0 t)^{n+1}}{(n+1)!} + \frac{(2\varepsilon^2 t_0 t)^{n+2}}{(n+2)!} + \cdots \right) \tag{41}$$

$$\leq e^{-\varepsilon^2 t^2 - \varepsilon^2 t_0^2} \frac{(2\varepsilon^2 t_0 t)^{n+1}}{(n+1)!} e^{2\varepsilon^2 t_0 t} \leq e^{2\varepsilon^2 D^2} \frac{(2\varepsilon^2 D^2)^{n+1}}{(n+1)!} \tag{42}$$

We invoke the basic estimate

$$n! > \left(\frac{n}{3}\right)^n \tag{43}$$

and obtain

$$\left| e^{-\varepsilon^2 (x-y)^2} - \sum_{n=0}^{r} \alpha_n(x-\bar{x})\beta_n(y-\bar{x}) \right| \leq e^{2\varepsilon^2 D^2} \frac{(2\varepsilon^2 D^2)^{n+1}}{\left(\frac{n+1}{3}\right)^{n+1}} = \left(\frac{6\varepsilon^2 D^2}{n+1}\right)^{n+1} e^{2\varepsilon^2 D^2} \tag{44}$$

Since we have eq. (38), which indicates

$$\frac{6\varepsilon^2 D^2}{n+1} < \frac{1}{2} \tag{45}$$

and therefore

$$\left| e^{-\varepsilon^2 (x-y)^2} - \sum_{n=0}^{r} \alpha_n(x - \bar{x})\beta_n(y - \bar{x}) \right| < \left(\frac{1}{2}\right)^{n+1} e^{2\varepsilon^2 D^2} < \delta \tag{46}$$

the last equation is due to the assumption eq. (38). $\qquad\square$

Remark 2. In practice, the estimate eq. (38) is quite conservative and smaller r can actually work very well.

The Crank-Nicolson discretization of eq. (25) on a uniform grid with spacing h and timestep Δt is

$$(I + \frac{1}{2}\Delta t A)u_j^{n+1} = (I - \frac{1}{2}\Delta t A)u_j^n \tag{47}$$

where

$$A = -a\delta_x^2 - b\delta_{2x} - c - \delta_L \tag{48}$$

here δ_x^2 and δ_{2x} are the standard second difference and central first difference. The discrete operator δ_L is defined through uniform grid Riemann sum of the truncated operator

$$(\delta_L u)_j = \sum_{j=-\infty}^{\infty} (u_{i+j} - u_i)\nu_j h = \sum_{j\in\mathcal{I}, j=-\infty}^{\infty} u_{i+j}\nu_j h - u_i\lambda h \quad \lambda = \sum_{j\neq 0, j=-\infty}^{\infty} \nu_j \tag{49}$$

Therefore, we have

$$A_{ij} = \begin{cases} \frac{2a}{h^2} - c + \lambda h & i = j \\ -\frac{a}{h^2} + \frac{b}{2h} - \nu_{-1}h & j = i - 1 \\ -\frac{a}{h^2} - \frac{b}{2h} - \nu_1 h & j = i + 1 \\ -\nu_{j-i}h & |j - i| \geq 2 \end{cases} \tag{50}$$

We consider the matrix $A_\pm = I \pm \frac{1}{2}A$. Note since the operator δ_x^2, δ_{2x} only contributes to the tridiagonal, any nonzero entry in $A_\pm$ in the off-diagonal more than one entry away from the diagonal must be $\mp\frac{1}{2}\nu_j h$ according to eq. (49).

We define the kernel associated with each Lévy measure by

$$K(x,y) = \nu(y - x) \tag{51}$$

then we have $A_{i'j'} = K(x_i, x_j)$ for $|i - j| \geq 2$, where i' and j' corresponds to x_i, x_j.

We carry out the stability analysis using the Fourier transform pair

$$u_j^n = \frac{1}{2\pi h} \int_{-\pi}^{\pi} \hat{u}^n(x) \exp(\mathrm{i}jx)dx \tag{52}$$

$$\hat{u}_j^n = h \sum_{-\infty}^{\infty} u_j^n \exp(-\mathrm{i}jx)dx \tag{53}$$

For simplicity, we assume that the domain is not truncated, i.e. $\mathcal{I} = \mathbb{Z}$. We need the following formula for

Lemma 2. Let

$$\eta_h(\theta) = \sum_{j=-\infty}^{\infty} (e^{\mathrm{i}jh\theta} - 1)v_j h \tag{54}$$

then we have

$$\delta_L e^{\mathrm{i}\theta x} = \eta_h(\theta)e^{\mathrm{i}\theta x} \tag{55}$$

where $\nu_j = \nu(jh)$. In the case ν_j is symmetric, we have

$$\eta_h(\theta) = -2 \sum_{j=-\infty}^{\infty} \sin^2\left(\frac{j\theta h}{2}\right) v_j h \leq 0 \tag{56}$$

8

Proof. By definition, we have

$$\delta_L e^{i\theta x} = \int_{-\infty}^{\infty} \left(e^{i(x+y)\theta} - e^{ix\theta} \right) \nu(y) dy = \eta_h(\xi) e^{i\theta x}$$

The other properties easily follow. $\square$

The Fourier transform of the numerical scheme gives

$$\hat{u}_i^{n+1} = \frac{1 - a\frac{\Delta t}{h^2}\sin^2\frac{\theta}{2} + \frac{b\Delta t}{2h}i\sin\theta + \frac{c\Delta t}{2} + \frac{\Delta t\eta_h(\theta)}{2}}{1 + a\frac{\Delta t}{h^2}\sin^2\frac{\theta}{2} - \frac{b\Delta t}{2h}i\sin\theta - \frac{c\Delta t}{2} - \frac{\Delta t\eta_h(\theta)}{2}} \hat{u}_i^n \tag{57}$$

Note we have

$$\left| 1 - a\frac{\Delta t}{h^2}\sin^2\frac{\theta}{2} + \frac{b\Delta t}{2h}i\sin\theta + \frac{c\Delta t}{2} + \frac{\Delta t\eta_h(\theta)}{2} \right|^2 =$$
$$\left| 1 - a\frac{\Delta t}{h^2}\sin^2\frac{\theta}{2} + \frac{c\Delta t}{2} + \frac{\Delta t\eta_h(\theta)}{2} \right|^2 + \left| \frac{b\Delta t}{2h}\sin\theta \right|^2 \tag{58}$$

$$\left| 1 + a\frac{\Delta t}{h^2}\sin^2\frac{\theta}{2} - \frac{b\Delta t}{2h}i\sin\theta - \frac{c\Delta t}{2} - \frac{\Delta t\eta_h(\theta)}{2} \right| =$$
$$\left| 1 + a\frac{\Delta t}{h^2}\sin^2\frac{\theta}{2} - \frac{c\Delta t}{2} - \frac{\Delta t\eta_h(\theta)}{2} \right|^2 + \left| \frac{b\Delta t}{2h}\sin\theta \right|^2 \tag{59}$$

Since we have $a \geq 0$, $c \leq 0$, $\eta_h(\theta) \leq 0$, we always have

$$\left| 1 + a\frac{\Delta t}{h^2}\sin^2\frac{\theta}{2} - \frac{c\Delta t}{2} - \frac{\Delta t\eta(\theta)}{2} \right| \geq \left| 1 - a\frac{\Delta t}{h^2}\sin^2\frac{\theta}{2} + \frac{c\Delta t}{2} + \frac{\Delta t\eta(\theta)}{2} \right| \tag{60}$$

Therefore, the model of the ratio in eq. (57) is always no greater than 1. Thus all the wave modes $e^{i\theta x}$ will not grow in magnitude if we carry out the Crank-Nicolson scheme.

The analytical solution to the model problem eq. (25) can be found using the Fourier transform when the coefficients are constant. For simplicity, we consider the case where the jump distribution is the Gaussian distribution, i.e.

$$\nu(y) = d\exp(-\varepsilon^2 y^2) \tag{61}$$

Then the corresponding characteristic function of the Lévy measure is

$$\eta(\xi) = -a\xi^2 + bi\xi + c + \int_{\mathbb{R}} (e^{i\xi y} - 1)\nu(y) dy \tag{62}$$

The last term can be computed analytically,

$$\int_{\mathbb{R}} (e^{i\xi y} - 1)\nu(y) dy = d\int_{\mathbb{R}} e^{i\xi y}\nu(y) dy - d\frac{\sqrt{\pi}}{\varepsilon} \tag{63}$$
$$= \frac{d\sqrt{\pi}}{\varepsilon} \left(e^{-\frac{\xi^2}{4\varepsilon^2}} - 1 \right) \tag{64}$$

Therefore, from

$$\hat{u}_t(\xi, t) = \eta(\xi)\hat{u}(\xi, 0) \tag{65}$$

9

we obtain
$$\hat{u}(\xi, t) = \exp(\eta(\xi)t)\hat{u}(\xi, 0) \tag{66}$$

The values of $u(x, t)$ at time $t = T$ can be evaluated through inverse Fourier transform

$$u(x_i, T) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{u}(\xi, T)e^{\mathrm{i}\xi x_i}d\xi = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(\eta(\xi)t)\hat{u}(\xi, 0)e^{\mathrm{i}\xi x_i}d\xi \tag{67}$$

which can be evaluated using FFT for uniform grids x_i.

Applications

In this section, we list several possible applications of the numerical scheme and fast algorithms. These applications are taken from literature which can be formulated as an integro-differential equation.

0.4   Option Pricing

Let S_t be the price of a financial asset which is modelled as a stochastic process under a martingale equivalence measure $\mathbb{Q}'$ and on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{Q})$.

One of the popular models is the exponential Lévy model which assumes

$$S_t = S_0 e^{rt + X_t} \tag{68}$$

where X_t is a Lévy process. Assume r is the interest rate. For a European call or put, the terminal payoff H_T at time T is associated with the underlying asset price S_T

$$H_T = H(S_T) \tag{69}$$

The value of the option is defined as a discounted conditional expection of H_T under the risk-adjusted martingale measure

$$C_t = \mathbb{E}[e^{-r(T-t)}H(S_T)|\mathcal{F}_t] = \mathbb{E}[e^{-r(T-t)}H(S_T)|S_t = S] \tag{70}$$

By introducing $\tau = T - t$, $x = \log\left(\frac{S}{S_0}\right)$, and define

$$u(x, \tau) = \mathbb{E}[h(x + Y_\tau)] \quad h(x) = H(S_0 e^x) \tag{71}$$

for sufficiently smooth u, by applying the Ito's formula for Lévy process we have the integro-differential equation

$$\frac{\partial u}{\partial \tau} = \frac{\sigma^2}{2} u_{xx} - (\sigma^2/2 - r + \alpha)u_x + \int_{\mathbb{R}} (u(x + y) - u(x) - u'(x)\mathbf{1}_{0 < |y| < 1}(y)y)\nu(y)dy \tag{72}$$

with initial condition

$$u(0, x) = h(x) \tag{73}$$

0.5 Quantum Mechanics

If the underlying stochastic process powering the random fluctuations is a Gaussian Brownian motion, we obtain the non relativistic Schrödinger's equation

$$i\hbar\partial_t\psi(x,t) = -\frac{\hbar^2}{2m}\partial_x^2\psi(x,t) \tag{74}$$

In recent years, there are a growing interest in non Gaussian stochastic process, and particularly the Lévy process. One of the popular models is the fractional quantum mechanics, where the stable processes are used as the underlying stochastic process. The popularity of the stable process is justified by the properties of scaling and self-similarity displayed by the process. For any distribution with power-law decay $\frac{1}{|x|^{1+\alpha}}$, $0 < \alpha < 1$ the generalized central limit theorem guarantee that their sum scaled by $\frac{1}{n^{1/\alpha}}$ converge to the α-stable distribution. If the variance are finite, i.e., $\alpha \geq 2$, then the central limit theorem holds, where their sum scaled by $\frac{1}{n^{1/2}}$, properly centered, and identically distributed, converge to the Gaussian distribution. This leads to the fractional Schrödinger equation

$$i\hbar\partial_t\psi(x,t) = D_\alpha(-\hbar^2\Delta)^{\alpha/2}\psi(x,t) \tag{75}$$

where $(-\hbar^2\Delta)^{\alpha/2}$ is the fractional Laplacian which can be defined through

$$(-\hbar^2\Delta)^{\alpha/2}\psi(x,t) = \frac{1}{(2\pi\hbar)^3}\int|\xi|^\alpha\hat{\psi}(\xi,t)\exp(i(\xi,x)/\hbar)d\xi \tag{76}$$

https://pdfs.semanticscholar.org/1dc0/9fedfa72c9f051da4642ff753d2bfe9f7a5f.pdf?_ga=2.23

More generally, other Lévy measure can be used to develop the quantum mechanics. The more general Schrödinger equation reads

$$i\hbar\partial_t\psi(x,t) = -\frac{\hbar^2}{2m}\partial_x^2\psi(x,t) - \hbar\int_\mathbb{R}[\psi(x+y,t)-\psi(x,t)]\nu(y)dy \tag{77}$$

https://arxiv.org/pdf/0805.0503.pdf
Some examples of the Lévy-Schrödinger equations are

- Relativistic.
$$i\hbar\partial_t\psi(x,t) = \sqrt{m^2c^4 - c^2\hbar^2\partial_x^2}\psi(x,t) \tag{78}$$

- Variance-Gamma laws

$$i\hbar\partial_t\psi(x,t) = -\frac{\lambda\hbar}{\tau}\int_\mathbb{R}\frac{\psi(x+y,t)-\psi(x,t)}{|y|}e^{-|y|/\hbar}dy \tag{79}$$

0.6 Turbulence Flow

https://arxiv.org/pdf/1803.05286.pdf
It is known that turbulence flow exhibits anomalous diffusion, i.e., the diffusion occurs over distance ξ may scale more than one half, $\xi \sim \mathcal{O}(t^{1/2})$. There are many efforts to model turbulence and capture these anomalies. One of the recent research is the the modeling of turbulence flow via the fractional Laplacian.

If we assume that the equilibrium probability distribution of particle speeds to be Lévy α-stable distributions instead of the Maxwell-Boltzmann distribution, we will arrive at the Navier-Stokes equation with the fractional Laplacian operator as a means to represent the mean friction force arising in a turbulence flow

$$\rho\frac{D\bar{u}}{Dt} = -\nabla p + \mu_\alpha \nabla^2 \bar{u} + \rho C_\alpha \int_{\mathbb{R}^3} \frac{\bar{u}(x',t) - \bar{u}(x,t)}{|x-x'|^{\alpha+3}} dx' \tag{80}$$

Extension

We now consider the general case where $\nu(y)$ can behave very badly. One such example is the fractional Laplacian where the Lévy measure is

$$\nu(y) = \frac{c_{1,s}}{|y|^{1+2s}} \tag{81}$$

where $s \in (0,1)$ and $c_{1,s}$ is a positive constant. Note in this case, $\int_{\mathbb{R}}(u(x+y) - u(x))\nu(y)dy$ must be understood in the principal value integration. The corresponding stochastic process associated with the fractional Laplacian is the α-stable process.

Consider the general singular integral operator

$$I(x) = \int_{\mathbb{R}} (u(x+y) - u(x) - \rho(y)u'(x)y)\nu(y)dy \tag{82}$$

where $\rho(y)u'(x)$ is a drift term to remove small activity from the jumps. $\rho(y)$ is a radial symmetric window function, satisfying

$$\begin{cases} 1 - \rho(y) \sim \mathcal{O}(y^4) & y \to 0 \\ \rho(y) = 0 & |y| \geq r \end{cases} \tag{83}$$

where $r > 0$ is a positive number.

As a reminder, we require $\nu(y)$ to satisfy the following conditions

$$\int_{-r}^{r} y^2 \nu(y)dy < \infty, \quad \int_{|y|\geq r} \nu(y)dy < \infty \tag{84}$$

The choice of $\rho(y)$ doesn't matter. In fact, if $\tilde{\rho}(x)$ is another window function that satisfies eq. (83), we have

$$\int_{\mathbb{R}} (u(x+y) - u(x) - \tilde{\rho}(y)u'(x)y)\nu(y)dy \tag{85}$$

$$= \int_{\mathbb{R}} (u(x+y) - u(x) - \rho(y)u'(x)y)\nu(y)dy + u'(x) \int_{\mathbb{R}} (\rho(y) - \tilde{\rho}(y)y)\nu(y)dy \tag{86}$$

we can add the second drift term into the model.

The first condition in eq. (83) is designed to take into consideration of the heavy tail case, where $\nu(y)$ can decay like $\mathcal{O}(1/|y|^{1+2s})$ for some $s \in (0,1)$. For example, in the special case

$$\nu(y) = c_{1,s} \frac{1}{|y|^{1+2s}} \tag{87}$$

12

we obtain the fractional Laplacian. Then $\int_{\mathbb{R}}(u(x+y)-u(x))\nu(y)dy$ is not well defined but only in the principle value integration, and it is easy to see

$$\text{p.v.} \int_{\mathbb{R}}(u(x+y)-u(x))\nu(y)dy = \int_{\mathbb{R}}(u(x+y)-u(x)-\rho(y)u'(x)y)\nu(y)dy \quad (88)$$

for any valid window function $\rho(y)$ thanks to the cancellation of the drift term due to symmetry.

Although $\int_{\mathbb{R}}(u(x+y)-u(x)-\rho(y)u'(x)y)\nu(y)dy$ is well-defined in this case, the integrand will behave like

$$(u(x+y)-u(x)-\rho(y)u'(x)y)\nu(y) = \mathcal{O}\left(\frac{1}{|y|^{2s-1}}\right) \quad (89)$$

in the case $s \to 0+$, we will have numerical difficulty if a direct numerical integration is applied. In the following, we will propose a numerical discretization for eq. (82) targeting at the most numerical challenging case described above

$$\nu(y) = \frac{n_0(y)}{|y|^{1+2s}} \quad (90)$$

where $n_0(y)$ is a bounded continuous function.

We make two assumptions on $u(x)$

- $u \in C(\mathbb{R})$

- Local smoothness. $u \in C^4([x-\delta, x+\delta])$ for some $\delta > 0$, i.e., u has fourth order derivative near the location where we want to evaluate $I(x)$.

- Far field asymptotic limit. The far field contribution

$$f_x(y) = \int_{\mathbb{R}} u(x+y)\nu(y)dy \quad (91)$$

is well defined. In the case $\frac{u(x+y)}{u(y)} \to f(y)$, this term can be approximated by $\int_{\mathbb{R}} f(y)dy$

The strategy is to singularity subtraction. We subtract a local diffusion term from eq. (82)

$$I(x) = \int_{\mathbb{R}}(u(x+y)-u(x)-\rho(y)u'(x)y-\frac{1}{2}\rho(y)u''(x)y^2)\nu(y)dy + \frac{1}{2}u''(x)\int_{\mathbb{R}}\rho(y)\nu(y)y^2dy \quad (92)$$

We can immediately split the first integral into two parts

$$I_1(x) = \int_{|y|\leq L_W}(u(x+y)-u(x)-\rho(y)u'(x)y-\frac{1}{2}\rho(y)u''(x)y^2)\nu(y)dy \quad (93)$$

and

$$I_2(x) = \int_{|y|>L_W}(u(x+y)-u(x)-\rho(y)u'(x)y-\frac{1}{2}\rho(y)u''(x)y^2)\nu(y)dy \quad (94)$$

$$= \int_{|x|>L_W}(u(x+y)-u(x))\nu(y)dy = f_x^{L_W} - u(x)\int_{|y|>L_W}\nu(y)dy \quad (95)$$

where L_W is a mask window and satisfies

$$L_W > r \qquad (96)$$

By Taylor expansion it is easy to see

$$u(x + y) - u(x) - \rho(y)u'(x)y - \frac{1}{2}\rho(y)u''(x)y^2 = \mathcal{O}(|y|^3) \qquad (97)$$

and therefore $I_1(x)$ will behave like $\mathcal{O}(|y|^{2-2s})$ near the origin. We will prove this fact in the appendix. That indicates the integrand becomes continuous near the origin. Thus $I_1(x)$ is well defined.

As $y \to \infty$, the term terms in $I_2(x)$ are both well defined according to the assumptions eqs. (84) and (91).

The second term in eq. (92)

$$I_3(x) = \frac{1}{2}u''(x) \int_{|y| \leq r} \nu(y)y^2 dy \qquad (98)$$

is a local diffusion term and the coefficient is well defined according to eq. (84).

We now focus on the numerical discretization of $I_1(x)$, $I_2(x)$ and $I_3(x)$. We divide the mask window into $2N$ uniform subintervals and consider the grid $\{ih : i \in \mathbb{Z}\}$, where $h = L_W/N$. We denote $u_i = u(ih)$.

Since the integrand in $I_1(x)$ is continuous, we can use a simple trapezoidal quadrature rule to approximate the integral. Assume the quadrature weights are w_j given by $w_{-N} = w_N = \frac{h}{2}$ and $w_j = h$, $j = -N+1, -N+2, \ldots, N-2, N-1$.

$$I_1(x_i) \approx \sum_{j=-N}^{N} u_{i+j}n(jh)w_j - u_i \sum_{j=-N}^{N} n(jh)w_j$$

$$- \frac{u_{i+1} - u_{i-1}}{h} \sum_{j=-N}^{N} \rho(jh)n(jh)jh - \frac{u_{i+1} + u_{i-1} - 2u_i}{2h^2} \sum_{j=-N}^{N} \rho(jh)(jh)^2 n(jh) \qquad (99)$$

where $\sum_{j=-N}^{N}$ denotes the summation excluding $j = 0$.

For $I_2(x)$, $f_x^{L_W}$ is either provided as an input or computed using a numerical quadrature and so is $\int_{|x|>L_W} \nu(y)dy$. We will see how these terms are obtained in the examples below.

$$I_2(x_i) \approx f_{x_i}^{L_W} - u(x_i) \int_{|y|>L_W} \nu(y)dy \qquad (100)$$

For I_3, a central difference scheme is applied to the second order derivative term.

$$I_3(x_i) \approx \frac{u_{i+1} + u_{i-1} - 2u_i}{2h^2} \int_{|y| \leqslant r} n(y)y^2 dy \qquad (101)$$

and the integral can either be computed analytically or numerically.

In practice, we want to compute $I(x)$ for $x \in [-L, L]$, according to table 1, we need to know the values of $u(x)$ on $[-L - L_W, L + L_W]$ and its corresponding far-field interactions. Figure 1 visualizes the relationship. To compute $I(x_i)$,

we need to compute the near field interaction and local interaction using values of $u(x)$ from the green area. The values of $u(x)$ are provided in the green and red area for computing $I(x)$, $x \in [-L, L]$.
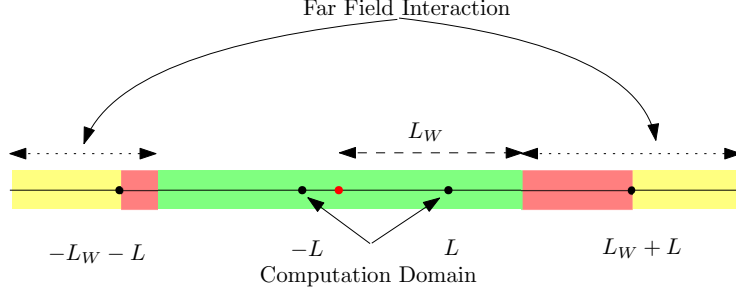

Figure 1: To compute $I(x_i)$, we need to compute the near field interaction and local interaction using values of $u(x)$ from the green area. The values of $u(x)$ are provided in the green and red area for computing $I(x)$, $x \in [-L, L]$.

Although we have used a different formula for the evaluation of the integral, we should soon realize that in the far-away off-diagonal parts, the entries are still $\nu_j h$ (except on the boundary), which the $\mathcal{H}$ matrix construction routine can still work.

Numerical Examples

0.7

Finally, we consider the case where $\nu(y)$ grows to infinity and has heavy tail. For simplicity, we will consider the specific case where $n(\mathbf{x}) = \frac{C_{d,s}}{|\mathbf{x}|^{d+2s}}$, i.e., the fractional Laplacian. Fortunately we can compute the analytical nonlocal derivative or gradient for some functions.

For the first example, we consider $u(x) = \exp(-x^2)$ in 1D. Then we have

$$(-\Delta)^s u(0) = 2^{2s}\Gamma\left(\frac{1+2s}{2}\right)/\sqrt{\pi} \qquad (102)$$

For this example, since $u(x)$ decay to zero exponentially, we can assume that the far-field interaction $f_x^{L_W} = 0$ is zero for sufficiently large L_W. The numerical value is computed using eqs. (99) to (101) and compared with the exact value eq. (102). The parameters are: $L_W = 5.0$, $r = 0.2$. The convergence plot is shown in fig. 2. We can see that the error converges like or better than $\mathcal{O}(h^2)$.
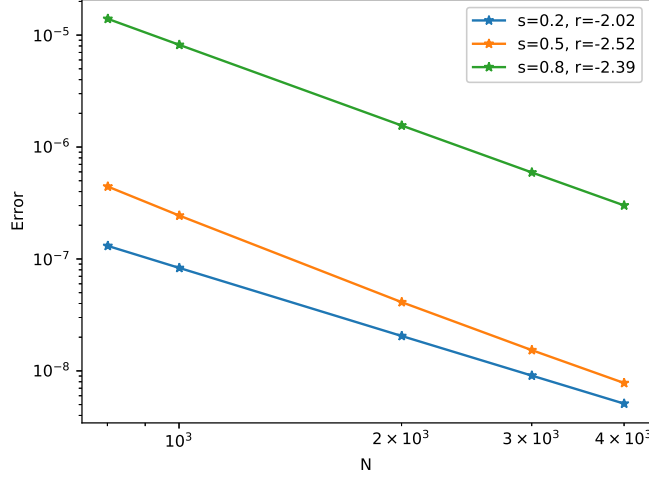
Figure 2: Numerical error for approximating eq. (102). The error converges like or better than $\mathcal{O}(h^2)$.

We also test the scheme on a challenging problem: the fractional Poisson problem. The PDE

$$\begin{cases} (-\Delta)^s u(x) = 1 & x \in [-1, 1] \\ u(x) = 0 & x \notin [-1, 1] \end{cases} \tag{103}$$

has a unique solution

$$u(x) = \frac{2^{2s}\Gamma(1+s)\Gamma\left(\frac{1+2s}{2}\right)}{\Gamma(1/2)}(1 - x^2)^s \tag{104}$$

Note that $u(x)$ is not smooth across the boundary. In fact, it only belongs to $C^{0,s}([-1, 1])$, the s-order Hölder space. Numerical algorithms usually exhibit reduced convergence. We use $L = 1.0$ and $Lw = 2.0$ so that the support of $u(x)$ is included in the near-field or local interaction. Thus we have $f_x^{Lw} = 0$. Since the current implementation only supports forward computation of the nonlocal operator, i.e., given function values, the nonlocal derivative or gradient is computed, we resort to a conjugate gradient approach for recovering $u(x)$ in $[-L, L]$.

Figure 3 presents the finite difference result obtained from our discretization. We can see that the convergence order is 1.0 or less, much worse than the Poisson problem where $\mathcal{O}(h^2)$ convergence rate is typical. We need to emphasize this is a universal problem faced by many fractional Laplacian models if a simple truncation trick is used.
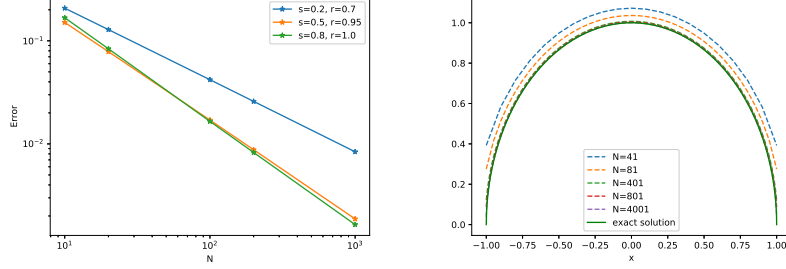
16

Figure 3: The finite difference result obtained from our discretization. The convergence order is 1.0 or less, much worse than the Poisson problem where $\mathcal{O}(h^2)$ convergence rate is typical

Finally, we also consider the computation of $(-\Delta)^u(\mathbf{x})$ in 2D, where

$$u(\mathbf{x}) = \frac{1}{2^{2s}\Gamma(1+s)^2}((1-|\mathbf{x}|^2)_+^s \tag{105}$$

The analytically result is known for $|\mathbf{x}| \leq 1$, which is

$$(-\Delta)^s u(\mathbf{x}) = 1 \quad |\mathbf{x}| \leq 1 \tag{106}$$

The numerical result is shown in fig. 4. Near the boundary, due to the nonsmoothness of $u(\mathbf{x})$, the algorithm is having a hard time computing the nonlocal gradient and therefore we see the oscillatory behavior. However, the computation for the region near the center is good, which does not suffer much from the far-away contribution from nonsmooth boundaries. In the center the error is only 4%. We used $L_W = 2.0$ and $L = 1.0$ in this case.
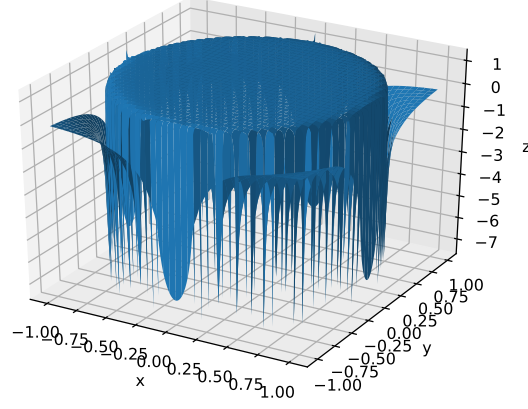
Figure 4: Numerical evaluation of eq. (106). Near the boundary, due to the nonsmoothness of $u(\mathbf{x})$, the algorithm is having a hard time computing the nonlocal gradient and therefore we see the oscillatory behavior. However, the computation for the region near the center is good, which does not suffer much from the far-away contribution from nonsmooth boundaries. In the center the error is only 4%.

These numerical examples demonstrates that the numerical scheme also works for $\nu(y)$ which has heavy tails.