Heat kernel and gradient estimates for kinetic SDEs with low regularity coefficients

P.E. Chaudru de Raynal[*], **S. Menozzi**[†], **A. Pesce**[‡], **X. Zhang**[§]

July 8, 2023

Abstract

We establish heat kernel and gradient estimates for the density of kinetic degenerate Kolmogorov stochastic differential equations. Our results are established under somehow minimal assumptions that guarantee the SDE is weakly well posed.

Keywords: degenerate Kolmogorov equations, kinetic dynamics, heat kernel and gradient estimates, parametrix

MSC: 60H10, 34F05

1 Introduction

1.1 Statement of the problem

We are interested in providing Aronson-like bounds and pointwise estimates for the full gradient of the transition probability density of the following *kinetic* system of SDEs:

$$\begin{cases} \mathrm{d}X_t^1 = F_1(t, X_t^1, X_t^2)\mathrm{d}t + \sigma(t, X_t^1, X_t^2)\mathrm{d}W_t, \\ \mathrm{d}X_t^2 = F_2(t, X_t^1, X_t^2)\mathrm{d}t, \end{cases} \tag{1.1}$$

where $(W_t)_{t\geqslant 0}$ stands for a d-dimensional Brownian motion on some stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ and for $i \in \{1, 2\}$, $t \geqslant 0$ the component X_t^i is $\mathbb{R}^d$-valued. This equivalently amounts to establish the announced estimates for the fundamental solution of the parabolic PDE associated with (1.1), which writes:

$$\begin{cases} \partial_s p\big(s, \boldsymbol{x}; t, \boldsymbol{y}\big) + \big\langle F_1(s, \boldsymbol{x}), \nabla_{x_1} p\big(s, \boldsymbol{x}; t, \boldsymbol{y}\big)\big\rangle + \big\langle F_2(s, \boldsymbol{x}), \nabla_{x_2} p\big(s, \boldsymbol{x}; t, \boldsymbol{y}\big)\big\rangle \\ \quad + \dfrac{1}{2}\mathrm{Tr}\Big(\sigma\sigma^*(s, \boldsymbol{x})\nabla_{x_1}^2 p\big(s, \boldsymbol{x}; t, \boldsymbol{y}\big)\Big) = 0, 0 \leqslant s < t,\ \boldsymbol{x} = (x_1, x_2), \boldsymbol{y} = (y_1, y_2) \in \mathbb{R}^{2d}, \\ p\big(s, \cdot; t, \boldsymbol{y}\big) \underset{s\uparrow t}{\longrightarrow} \delta_{\boldsymbol{y}}(\cdot). \end{cases} \tag{1.2}$$

[*]Laboratoire de Mathématiques Jean Leray, 2, rue de la Houssinière BP 92208 F-44322 Nantes Cedex 3. Paul-Eric.Deraynal@univ-nantes.fr

[†]Laboratoire de Modélisation Mathématique d'Evry (LaMME), Université d'Evry Val d'Essonne, 23 Boulevard de France 91037 Evry, France and Laboratory of Stochastic Analysis, HSE, Pokrovsky Blvd, 11, Moscow, Russian Federation. stephane.menozzi@univ-evry.fr

[‡]Dipartimento di Matematica, Piazza di Porta San Donato, 5 Bologna (Italy), antonello.pesce2@unibo.it

[§]School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, P.R. China, Email: XichengZhang@gmail.com

Importantly, we aim at obtaining such estimates under somehow *minimal* assumptions, relying on *minimal* conditions required on the SDE to be well posed, in a weak sense. For our approach to work, we will assume a kind of *weak* Hörmander condition - the Jacobian $(\nabla_{x_1} F_2)$ has full rank and the diffusion coefficient σ is bounded and separated from zero - however the coefficients can be rather rough in their entries, namely, measurable with respect to the time variable and belonging to suitable *anisotropic Hölder spaces* in the spatial variables. In particular, we emphasize that the drift term F can be unbounded in all its variables and entries. Through the analysis, some thresholds for the Hölder regularity of the drift with respect to the second (whence degenerate) variable x_2 will appear. Such thresholds are related to the degenerate nature of the system of interest and appear to be rather sharp as they are precisely the ones which provide a sufficient and (almost) necessary condition for the system to be well posed, see [6].

From now on we shall use bold letters $\mathbf{X}$ and $\mathbf{F}$ to denote vectors (X^1, X^2) and (F_1, F_2) in $\mathbb{R}^{2d}$. Let $B = (\mathbb{I}_{d\times d}, 0_{d\times d})^*$ be a $2d \times d$-matrix, where $*$ stands for the transpose. Using these notations, we can rewrite SDE (1.1) in the following compact form:

$$\mathrm{d}\mathbf{X}_t = \mathbf{F}(t, \mathbf{X}_t)\mathrm{d}t + B\sigma(t, \mathbf{X}_t)\mathrm{d}W_t. \tag{1.3}$$

Related applications: These kinds of *kinetic* (or speed/position) systems appear in several application fields. For instance (1.1) describes the dynamics of some Hamiltonian systems. For a Hamilton function $H(\boldsymbol{x}) = V(x_2)+|x_1|^2/2$, where V is a potential and $|x_1|^2/2$ corresponds to the kinetic energy of a particle with unit mass, the corresponding drift $\mathbf{F}_H$ would write $\mathbf{F}_H(\boldsymbol{x}) = (-\nabla_{x_2} V(x_2), x_1)^*$. Adding a damping term $\mathbf{D}(\boldsymbol{x})$, i.e. for $\mathbf{F}(\boldsymbol{x}) = (\mathbf{F}_H - \mathbf{D})(\boldsymbol{x})$, leads to investigate the long time behavior of the system, we can e.g. refer to the works [13], [15] for related discussions, to the monograph [36] for applications in mechanics or to [39], [27] for numerical approximations of the invariant measures.

In mathematical finance, equation (1.1) can be related to the model used to price path-dependent contracts, such as Asian options (see, [1] or [8] for recent developments).

We choose here to focus on the very object behind, the density, over a finite time interval. To expose some of the particular features of the model, let us start our discussion with the (striking) Gaussian setting.

Gaussian case and the Hörmander condition: For illustrative purposes let us examine the case $F_1 \equiv 0$, $\sigma \equiv 1$ and $F_2(t, X^1_t, X^2_t) = X^1_t$, which corresponds to the Langevin dynamics in its simplest form:

$$\mathrm{d}X^1_t = \mathrm{d}W_t, \quad \mathrm{d}X^2_t = X^1_t\mathrm{d}t, \qquad t \geqslant 0, \tag{1.4}$$

which equivalently rewrites in the short form (1.3)

$$\mathrm{d}\mathbf{X}_t = \mathbf{A}\mathbf{X}_t\mathrm{d}t + B\mathrm{d}W_t, \mathbf{A} = \begin{pmatrix} 0_{d\times d} & 0_{d\times d} \\ \mathbb{I}_{d\times d} & 0_{d\times d} \end{pmatrix}, \qquad t \geqslant 0. \tag{1.5}$$

In his seminal work [19], Kolmogorov derived the fundamental solution for the PDE (1.2) associated with the above process. For an initial value $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^{2d}$, we have for $t \geqslant 0$,

$$\mathbf{X}_t = (X_t^1, X_t^2) = \left(x_1 + W_t, x_2 + x_1 t + \int_0^t W_s \mathrm{d}s\right) = \exp(\mathbf{A}t)\boldsymbol{x} + \int_0^t \exp\big(\mathbf{A}(t-s)\big)B\mathrm{d}W_s,$$

which is a Gaussian process with mean and covariance matrix respectively given by

$$\boldsymbol{\theta}_t(\boldsymbol{x}) = \exp(\mathbf{A}t)\boldsymbol{x} = (x_1, x_2 + x_1 t), \qquad \mathbf{K}_t = \begin{pmatrix} t\mathbb{I}_{d\times d} & \frac{t^2}{2}\mathbb{I}_{d\times d} \\ \frac{t^2}{2}\mathbb{I}_{d\times d} & \frac{t^3}{3}\mathbb{I}_{d\times d} \end{pmatrix}. \tag{1.6}$$

The matrix $\mathbf{K}_t$ is positive definite for every $t > 0$ and therefore the process admits a density for every $t > 0$, explicitly given by

$$\boldsymbol{y} \mapsto \Big(\frac{\sqrt{3}}{\pi t^2}\Big)^d \exp\left(-\tfrac{1}{2}|\mathbf{K}_t^{-\frac{1}{2}}(\boldsymbol{y} - \boldsymbol{\theta}_t(\boldsymbol{x}))|^2\right) = p(0, \boldsymbol{x}; t, \boldsymbol{y}) =: p(\boldsymbol{x}; t, \boldsymbol{y}). \tag{1.7}$$

In particular, there exist constants $0 < c_- < c_+$ such that

$$\Big(\frac{\sqrt{3}}{\lambda\pi t^2}\Big)^d \exp\big(-c_+|\mathbb{T}_t^{-1}(\boldsymbol{y} - \boldsymbol{\theta}_t(\boldsymbol{x}))|^2\big) \leqslant p(\boldsymbol{x}; t, \boldsymbol{y}) \leqslant \Big(\frac{\sqrt{3}}{\lambda\pi t^2}\Big)^d \exp\big(-c_-|\mathbb{T}_t^{-1}(\boldsymbol{y} - \boldsymbol{\theta}_t(\boldsymbol{x}))|^2\big), \tag{1.8}$$

where, for $t > 0$,

$$\mathbb{T}_t = \begin{pmatrix} t^{\frac{1}{2}}\mathbb{I}_{d\times d} & 0_{d\times d} \\ 0_{d\times d} & t^{\frac{3}{2}}\mathbb{I}_{d\times d} \end{pmatrix}, \tag{1.9}$$

is the scale matrix which precisely reflects the multi-scale behavior of the components.

Let us first remark that, in Hörmander form, the generator of the process in (1.4) writes

$$L = A_1^2 + A_0,\ A_1 = \nabla_{x_1},\ A_0(\boldsymbol{x}) := x_1 \nabla_{x_2}, \tag{1.10}$$

so that, denoting by $[\cdot,\cdot]$ the Lie bracket, $[A_1, A_0] = \nabla_{x_2}$ and $\mathrm{Span}\{A_1, [A_1, A_0]\} = \mathbb{R}^{2d}$. Importantly, we see that the drift is really needed to span the whole space. This is why we speak about *weak* Hörmander condition. As we have seen, this kind of assumption leads to a multi-scale behavior of the heat-kernel as opposed to the strong Hörmander condition, i.e. when the diffusive vector fields and their Lie brackets span the space. In that case, two-sided heat kernel bounds, which exhibit a usual parabolic scaling, in $\sqrt{t}$ w.r.t. the Carnot metric induced by the vector fields, are available in [21]. There is therefore a drastic difference between these two types of assumptions.

Let us eventually mention that, since we are going to consider *rough* coefficients, we will not be able to perform Lie bracketing to justify the existence of the density from the Hörmander condition. The non degeneracy of $\nabla_{x_1} F_2$ can somehow be seen as a *mild weak* Hörmander type condition[1]. We emphasize that it is precisely this term which makes the Kolmogorov example work, because it precisely allows the noise on the first component to propagate to the second one.

[1]In the document, we will further refer with a slight terminology abuse to this assumption as the weak type Hörmander condition.

Secondly, it can be observed from this example that the time-scale in the density is not diffusive: the fluctuations of the two components X^1, X^2 are of order of $t^{1/2}$ and $t^{3/2}$ respectively, which corresponds to the intuition that the typical time-scale of an integrated Brownian motion should be equal to the integral of the time-scale of the Brownian motion. This phenomenon also appears in the deviation term in the exponential. The growth rate is different for the two components. Also, the unbounded drift term induces deviations w.r.t. to the transport of the initial condition by the underlying deterministic differential system $\boldsymbol{\theta}_t(\boldsymbol{x})$ and not the starting point itself $\boldsymbol{x}$, normalized w.r.t. the previous intrinsic time scales.

Similarly, it is seen that there exists $C \geqslant 1$ s.t. for $i \in \{1, 2\}$,

$$\begin{aligned} |\nabla_{x_i} p(\boldsymbol{x}; t, \boldsymbol{y})| &\leqslant \left|\left((\mathbf{K}_t^{-\frac{1}{2}} \nabla \boldsymbol{\theta}(\boldsymbol{x}))^* \mathbf{K}_t^{-\frac{1}{2}} (\boldsymbol{\theta}_t(\boldsymbol{x}) - \boldsymbol{y})\right)\right)_i |p(\boldsymbol{x}; t, \boldsymbol{y}) \\ &\leqslant \frac{C}{t^{\frac{2i-1}{2}}} \left(\frac{\sqrt{3}}{\pi t^2}\right)^d \exp\left(-C^{-1} |\mathbb{T}_t^{-1}(\boldsymbol{y} - \boldsymbol{\theta}_t(\boldsymbol{x}))|^2\right). \end{aligned}$$

Namely, a differentiation with respect to the non-degenerate variable induces an additional time singularity of the corresponding typical order rate $t^{-1/2}$ whereas a differentiation with respect to the degenerate variable induces a time singularity in $t^{-3/2}$ i.e. at its corresponding typical rate.

Our goal in the current work is to extend those heat kernel and gradient bounds, obtained directly for the Kolmogorov example, to the densities of SDEs solving (1.1) under somehow minimal smoothness assumptions on the coefficients.

Available results and *minimal* conditions from a regularization by noise perspective. Since the seminal work of Kolmogorov [19], such equations have been thoroughly investigated in the literature both from the analytic or probabilistic viewpoint[2]. Existence of fundamental solution for the underlying parabolic PDE (1.2) was first obtained through a *parametrix* type perturbation technique, for smooth enough coefficients, by Weber [41]. We can also refer to Sonin for further results in that direction [37]. On the other hand, density estimates were derived in [20] (global upper and lower *diagonal* bound) and then extended to more general models of SDEs that can be seen as perturbed ODEs for which a noise acting on the first component will transmit to the whole chain of ODEs through a weak Hörmander like condition. For such models we refer to [9], [29]. From those works one can derive two-sided heat kernel bounds for the density of (1.1) when the drift is globally Lipschitz in space, i.e. when the drift part of the dynamics in (1.1) can be associated with a usual well-posed ODE, and when the diffusion coefficient is Hölder continuous in space. Eventually, in a smooth framework, Pigato derived in [35] heat kernel and gradient estimates for the system as well as short time asymptotics. Except for the diffusion coefficient, the aforementioned works do not take advantage of the propagation of the noise through the system, in the sense that the drift is always assumed to be (at least) a Lipschitz in space function. As already claimed, our aim consists in deriving those bounds under rather *minimal* conditions, meaning that we manage to benefit from the regularization by noise phenomenon, see the Saint Flour lectures notes [11] for an overview of such kind of phenomenom.

Regularization by noise for degenerate system was investigated in e.g. [3] from a strong point of view, meaning that the author exhibited therein, within the framework of Hölder spaces, some mini-

[2]Let us also recall that the work [19] was recalled by Hörmander as the starting point of his general theory of hypoellipticity [16].

mal thresholds on F that guarantee strong well posedness of equation (1.1), in spite of a non Lipschitz drift. Generalization to Hölder-Dini coefficients were investigated in [40] with other techniques but somehow similar threholds. For rougher drifts, corresponding namely to Bessel potentials, but for $F_2(t,x) = x_1$, the strong well posedness was also derived (for the same previous *regularity threshold on the degenerate variable*) in [10] and [43] where the critical case for the regularity index is considered. However, it is well known that existence of the density does not rely on strong well-posedness of the system, but more generally on its weak well-posedness.

In the current degenerate setting, the weak regularization by noise was investigated for the kinetic case in [4]. The author derived therein, still in the setting of Hölder spaces, smaller thresholds on the regularity index of the drift w.r.t. the degenerate variable that yield weak uniqueness. It is also importantly shown that these thresholds are (almost) sharp in the sense that, when the drift of the degenerate component has Hölder regularity below the threshold, there are counter-examples to weak uniqueness. This weak well-posedness and associated counter-examples have then been extended to a full chain of perturbed ODEs in [6]. In this latter work, the Authors exhibit an (almost) sharp characterization, at each level of the ODE chain, of the regularity index needed in each variable to restore weak well-posedness. In comparison with our current setting, it is proved therein that the drift of each component feels differently the degenerate variable: whereas one only needs positive regularity indexes for the drift of the non degenerate component w.r.t. all variables for weak uniqueness to hold, a threshold of $1/3$ appears for the regularity index of the drift of the degenerate component w.r.t. to the degenerate variable. We will try to match this setting as much as possible for what we aim at doing here. We refer to Theorem 1.1 and 1.4 and associated remarks for details.
Again, having a good understanding of the density under minimal conditions on the coefficients can be very useful in connection with some related non-linear equations. We can e.g. refer to Section 1.3 in [18] where a *toy* non-linear model, which shares some properties with the Landau equation, involving an average of the density w.r.t. the velocity variable is considered as diffusion coefficient.
Objective and strategy. Many other type of estimates have been established for the SDE (1.1) or its formal generator. Let us mention among them: Harnack inequalities [24], [34]; related heat kernel estimates for operators in divergence form with measurable coefficients [22], [23]; Schauder estimates, see [18] for the current framework and [5] for an even more general case (one can also refer to [14] for an extension to kinetic non-local operators - with an application to strong well posedness - and to [26] as well for a more general framework); L^p estimates, see [17], [30] or [7]. Let us eventually mention the work [33] which deals with the associated Stochastic PDE in the two-dimensional case or [44] which investigates the well-posedness of a McKean-Vlasov version of (1.2) through the De Giorgi approach.

We will focus here on the density/heat kernel and will adapt the approach already considered in [31], [32], for non degenerate SDEs with unbounded drift respectively Brownian and stable driven, to the current degenerate case. The first step consists in obtaining two-sided estimates. This is done using *forward* type parametrix or Duhamel type expansions, as e.g. considered in the classic non-degenerate case in [12] or [28], with the additional difficulty that, because of the unbounded drift the parametrix series needs to be truncated. The tails of the series are controlled through stochastic control arguments (see [9], [42]). For the estimates on the derivatives the idea consists in mixing *forward* and *backward* Duhamel expansions and to consider suitable normalizations which exploit thoroughly the underlying two-sided estimates.

We restrict in this work to the kinetic case for simplicity. We believe that our main results would extend to the full perturbed chain of ODE as considered in [6] under suitable assumptions on the coefficients. In this more general setting the idea would be to couple the current approach with the computations performed in [2] to derive strong well posedness for the full chain.

The article is organized as follows. We state our main results (Theorems 1.1 and 1.4) in Section 1.2. Section 2 gathers some technical results about mollified flows associated with Hölder in space coefficients and also addresses the corresponding deterministic control problem which will be useful for the two-sided heat kernel estimate. We will establish in Section 3 our main results for smooth coefficients which satisfy the assumptions of Theorems 1.1 and 1.4 and we will carefully prove that the constants in the estimates obtained do not depend on such smoothness. This is precisely why we then derive our main results through compactness arguments detailed in Section 4.

1.2 Statement of main results

Let $d, l \in \mathbb{N}$. For $j \in \{0\} \cup \mathbb{N}$ and $\gamma \in [0,1)$, let $\mathscr{C}^{j+\gamma}(\mathbb{R}^d;\mathbb{R}^l)$ be the space of Hölder functions from $\mathbb{R}^d$ to $\mathbb{R}^l$ defined by

$$\mathscr{C}^{j+\gamma}(\mathbb{R}^d;\mathbb{R}^l) := \left\{ f : \|f\|_{\mathscr{C}^{j+\gamma}(\mathbb{R}^d;\mathbb{R}^l)} := \sum_{k=1}^{j} \|\nabla^k f\|_{L^\infty(\mathbb{R}^d;\mathbb{R}^l)} + \sup_{x\neq y, |x-y|\leqslant 1} \frac{|\nabla^j f(x) - \nabla^j f(y)|}{|x-y|^\gamma} < \infty \right\},$$

where ∇^k stands for the k-order gradient. Note that the functions in $\mathscr{C}^{j+\gamma}(\mathbb{R}^d;\mathbb{R}^l)$ can be unbounded and have sublinear growth.

Importantly the functions in $\mathscr{C}^0(\mathbb{R}^d;\mathbb{R}^l)$ can be possibly discontinuous and also satisfy (see [40], Lemma 2.3)

$$\sup_{x\neq y} \frac{|f(x)-f(y)|}{1+|x-y|} < +\infty.$$

We denote by $\boldsymbol{x} = (x_1, x_2)$ a point in $\mathbb{R}^d \times \mathbb{R}^d$ and by ∇_{x_1}, ∇_{x_2} the gradients with respect to the first and second set of variables, respectively. Following the previous discussion, it is natural to endow $\mathbb{R}^{2d}$ with an anisotropic distance, corresponding to the intrinsic scale matrix (1.9):

$$|\boldsymbol{x}|_{\mathbf{d}} := |x_1| + |x_2|^{\frac{1}{3}}, \qquad \boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^{2d}. \tag{1.11}$$

Next, we recall the definition of a *anisotropic Hölder spaces* associated with (1.11) (see, for instance [25], [5]). We say that a vector valued function $f \in \mathscr{C}^{j+\gamma}_{\mathbf{d}}(\mathbb{R}^{2d};\mathbb{R}^l)$ if

$$\|f\|_{\mathscr{C}^{j+\gamma}_{\mathbf{d}}(\mathbb{R}^{2d};\mathbb{R}^l)} := \sup_{x_2\in\mathbb{R}^d} \|f(\cdot,x_2)\|_{\mathscr{C}^{j+\gamma}(\mathbb{R}^d;\mathbb{R}^l)} + \sup_{x_1\in\mathbb{R}^d} \|f(x_1,\cdot)\|_{\mathscr{C}^{(j+\gamma)/3}(\mathbb{R}^d;\mathbb{R}^l)} < \infty. \tag{1.12}$$

In particular, for $f \in \mathscr{C}^{1+\gamma}_{\mathbf{d}}(\mathbb{R}^{2d};\mathbb{R})$, by Taylor's expansion, we have

$$|\mathcal{T}_f(\boldsymbol{x},\boldsymbol{y}) := f(\boldsymbol{x}) - f(\boldsymbol{y}) - \nabla_{x_1} f(\boldsymbol{y})(\boldsymbol{x}-\boldsymbol{y})_1| \leqslant C_\gamma \|f\|_{\mathscr{C}^{1+\gamma}_{\mathbf{d}}} |\boldsymbol{x}-\boldsymbol{y}|^{1+\gamma}_{\mathbf{d}}. \tag{1.13}$$

We assume the following conditions to hold:

($\mathbf{H}_\sigma^\gamma$) There exist $\gamma \in (0,1]$ and $\kappa_0 \geqslant 1$ such that for all $(t, \boldsymbol{x}) \in \mathbb{R}_+ \times \mathbb{R}^{2d}$ and $\xi \in \mathbb{R}^d$,

$$\kappa_0^{-1}|\xi|^2 \leqslant \langle \sigma\sigma^*(t,\boldsymbol{x})\xi, \xi\rangle \leqslant \kappa_0|\xi|^2$$

and

$$|\sigma(t,\boldsymbol{x}) - \sigma(t,\boldsymbol{y})| \leqslant \kappa_0|\boldsymbol{x}-\boldsymbol{y}|_{\mathbf{d}}^\gamma,$$

where $|\cdot|$ denotes the Euclidean norm, $\langle\cdot,\cdot\rangle$ is the inner product and * stands for the transpose.

($\mathbf{H}_\mathbf{F}^\gamma$) For some $\gamma \in (0,1]$ and $\kappa_1, \kappa_2 > 0$, it holds that

$$|F_i(t,\mathbf{0})| \leqslant \kappa_i,\ i = 1,2,\ \|F_1(t,\cdot)\|_{\mathscr{C}_\mathbf{d}^0} \leqslant \kappa_1,\ \|F_2(t,\cdot)\|_{\mathscr{C}_\mathbf{d}^{1+\gamma}} \leqslant \kappa_2.$$

Moreover, there exists a closed convex subset $\mathcal{E} \subset GL_d(\mathbb{R})$ (the set of all invertible $d \times d$ matrices) such that $\nabla_{x_1}F_2(t,\boldsymbol{x}) \in \mathcal{E}$ for all $t \geqslant 0$ and $\boldsymbol{x} \in \mathbb{R}^{2d}$.

We introduce the following notation

$$g_\lambda(t,\boldsymbol{x}) := t^{-2d}\mathrm{e}^{-|\mathbb{T}_t^{-1}\boldsymbol{x}|^2/(2\lambda)} \tag{1.14}$$

as well as the following set of parameters for later use: for $T > 0$,

$$\Theta_T := (T, \kappa_0, \kappa_1, \kappa_2, d, \gamma, \mathcal{E}).$$

Eventually, to state our main results we need to introduce a *mollified* flow associated with the drift F in (1.1), which under **($\mathbf{H}_\mathbf{F}^\gamma$)** is *rough*. Namely,

$$\dot{\widetilde{\boldsymbol{\theta}}}_{t,s}(\boldsymbol{x}) = \widetilde{\mathbf{F}}(t, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x})), \quad \widetilde{\boldsymbol{\theta}}_{s,s}(\boldsymbol{x}) = \boldsymbol{x}, \tag{1.15}$$

where

$$\widetilde{\mathbf{F}}(t,\boldsymbol{x}) = \Big([F_1(t,\cdot) * \rho_1](\boldsymbol{x}), [F_2(t,\cdot) * \rho_{|t-s|^{3/2}}](\boldsymbol{x})\Big),$$

and $\rho_\varepsilon(\boldsymbol{x}) = \varepsilon^{-2d}\rho(\varepsilon^{-1}\boldsymbol{x})$ and ρ is a smooth density function with compact support and $*$ stands for the convolution in space. The first regularization, performed at a macro level, is very natural to introduce a flow since the initial drift coefficient is not necessarily smooth. Regularizing the second component, allows as well to have a flow defined in the classical sense. The regularization parameter, corresponding to the intrinsic time scale of the component allows to have the *equivalence* between this flow and any other regularized flow (see Remark 1.3 and Lemma 4.1 for details).

Our first main result of this paper is stated as follows.

Theorem 1.1. *Under* **($\mathbf{H}_\sigma^\gamma$)** *and* **($\mathbf{H}_\mathbf{F}^\gamma$)** *with $\gamma \in (0,1]$, for any $T > 0$ and $0 \leqslant s < t \leqslant T$, there exists a unique weak solution $\mathbf{X}_{t,s}(\boldsymbol{x})$ of* (1.1) *starting from $\boldsymbol{x}$ at time s which admits a density $p(s,\boldsymbol{x};t,\boldsymbol{y})$ continuous in $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^{2d}$. Moreover, $p(s,\boldsymbol{x};t,\boldsymbol{y})$ enjoys the following estimates:*

(i) (Two sided estimates) There are $\lambda_0, C_0 \geqslant 1$ depending on Θ_T such that for all $0 \leqslant s < t \leqslant T$ and $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^{2d}$,

$$C_0^{-1}g_{\lambda_0^{-1}}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big) \leqslant p(s,\boldsymbol{x};t,\boldsymbol{y}) \leqslant C_0 g_{\lambda_0}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big). \tag{1.16}$$

(ii) (Gradient estimate in x_1) There exist constants λ_1, $C_1 \geqslant 1$ depending on Θ_T such that for any $0 \leqslant s < t \leqslant T$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2d}$,

$$|\nabla_{x_1} p(s, \boldsymbol{x}; t, \boldsymbol{y})| \leqslant C_1 (t-s)^{-\frac{1}{2}} g_{\lambda_1}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big). \tag{1.17}$$

(iii) (Hölder estimate in $\boldsymbol{x}$) Let $\eta_0, \eta_1 \in (0,1)$. For $j = 0,1$, there exist constants λ_j, $C_j \geqslant 1$ depending on Θ_T and η_j such that for any $0 \leqslant s < t \leqslant T$ and $\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{y} \in \mathbb{R}^{2d}$,

$$\begin{aligned}\big|\nabla^j_{x_1} p(s, \boldsymbol{x}; t, \boldsymbol{y}) - \nabla^j_{x_1} p(s, \boldsymbol{x}'; t, \boldsymbol{y})\big| &\leqslant C_j |\boldsymbol{x} - \boldsymbol{x}'|^{\eta_j}_{\mathbf{d}} (t-s)^{-(\frac{j}{2}+\eta_j)} \\ &\times \Big(g_{\lambda_j}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big) + g_{\lambda_j}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}') - \boldsymbol{y}\big)\Big).\end{aligned} \tag{1.18}$$

Remark 1.2. Under $(\mathbf{H}^\gamma_\mathbf{F})$, F_1 may be unbounded and discontinuous. For instance, if $F_1(\boldsymbol{x}) = F_{11}(\boldsymbol{x}) + F_{12}(\boldsymbol{x})$ and $F_2(x_1, x_2) = x_1$, where F_{11} is bounded measurable and F_{12} is global Lipschitz, then $\mathbf{F} = (F_1, F_2)$ satisfies $(\mathbf{H}^1_\mathbf{F})$. This example corresponds to the standard kinetic SDEs.

Let us emphasize that this Theorem holds under the assumptions that have been shown in [6] to be minimal to guarantee weak uniqueness for the solution of (1.1). In particular the two sided estimates (1.16) specify the Krylov type estimate of [6] which roughly said that in $L^q - L^p$ norms (for suitable indexes p, q) the density behaved *as* the Kolmogorov one appearing in (1.16).

Remark 1.3 (About the flow in the above estimates)**.** We point out that the above estimates could also be stated replacing $\widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x})$ introduced in (1.15) by any Peano flow $\boldsymbol{\theta}^{(1)}_{t,s}(\boldsymbol{x})$ solving:

$$\dot{\boldsymbol{\theta}}^{(1)}_{t,s}(\boldsymbol{x}) = \mathbf{F}^{(1)}(t, \boldsymbol{\theta}^{(1)}_{t,s}(\boldsymbol{x})), \quad \boldsymbol{\theta}^{(1)}_{s,s}(\boldsymbol{x}) = \boldsymbol{x},$$

where

$$\mathbf{F}^{(1)}(t, \boldsymbol{x}) = \Big([F_1(t,\cdot) * \rho_1](\boldsymbol{x}), F_2(t,\cdot)(\boldsymbol{x})\Big).$$

Pay attention that F_2 is not regularized whereas the possibly discontinuous component F_1 still needs to be to define an underlying flow.

Indeed it can be shown, similarly to the estimates established in Lemma 4.1 below, that there exists a constant $C := C(\Theta_T)$ s.t.:

$$\big|\mathbb{T}^{-1}_{t-s}\big(\boldsymbol{\theta}^{(1)}_{t,s}(\boldsymbol{x}) - \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x})\big)\big| \leqslant C,$$

i.e. $\boldsymbol{\theta}^{(1)}_{t,s}(\boldsymbol{x})$ and $\widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x})$ are equivalent with respect to the intrinsic scales.

In the above theorem, the regularity assumptions on σ and $\mathbf{F}$ are almost sharp. To obtain the second order derivative estimate in x_1 and the first order gradient estimate in x_2, we have to make further regularity assumptions as stated below.

Theorem 1.4. *In the situation of Theorem 1.1, we also assume that for the same $\gamma \in (0,1]$,*

$$\|F_1(t,\cdot)\|_{\mathscr{C}^\gamma_\mathbf{d}} \leqslant \kappa_1, \quad t \geqslant 0. \tag{1.19}$$

(i) (Second order derivative estimate in x_1) There exist constants $\lambda_1,\ C_1 \geqslant 1$ depending on Θ_T such that for any $0 \leqslant s < t \leqslant T$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2d}$,

$$\big|\nabla^2_{x_1} p(s, \boldsymbol{x}; t, \boldsymbol{y})\big| \leqslant C_1 (t-s)^{-1} g_{\lambda_1}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big). \tag{1.20}$$

(ii) (Hölder estimate of $\nabla^2_{x_1} p$ in $\boldsymbol{x}$) For any $\eta_2 \in (0, \gamma)$, there exist constants $\lambda_2,\ C_2 \geqslant 1$ depending on Θ_T and η_2 such that for any $0 \leqslant s < t \leqslant T$ and $\boldsymbol{x}, \boldsymbol{x}'\boldsymbol{y} \in \mathbb{R}^{2d}$,

$$\begin{aligned} \big|\nabla^2_{x_1} p(s, \boldsymbol{x}; t, \boldsymbol{y}) - \nabla^2_{x_1} p(s, \boldsymbol{x}'; t, \boldsymbol{y})\big| &\leqslant C_2 |\boldsymbol{x} - \boldsymbol{x}'|^{\eta_2}_{\mathbf{d}} (t-s)^{-1-\frac{\eta_2}{2}} \\ &\times \Big(g_{\lambda_2}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big) + g_{\lambda_2}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}') - \boldsymbol{y}\big) \Big). \end{aligned} \tag{1.21}$$

If σ also satisfies that

$$|\sigma(t, \boldsymbol{x}) - \sigma(t, \boldsymbol{y})| \leqslant \kappa_0 (|(\boldsymbol{x} - \boldsymbol{y})_1|^{\alpha} + (\boldsymbol{x} - \boldsymbol{y})_2|^{\frac{1+\gamma}{3}}), \tag{1.22}$$

where γ is the same as in Theorem 1.1 and $\alpha \in ((1-\gamma) \vee \gamma, 1]$, and

$$|F_1(t, \boldsymbol{x}) - F_1(t, \boldsymbol{y})| \leqslant \kappa_1 (|(\boldsymbol{x} - \boldsymbol{y})_1|^{\gamma} + (\boldsymbol{x} - \boldsymbol{y})_2|^{\frac{1+\gamma}{3}}), \tag{1.23}$$

then

(iii) (Gradient estimate in x_2) there exist $\lambda_3,\ C_3 \geqslant 1$ depending on Θ_T and α such that for any $0 \leqslant s < t \leqslant T$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2d}$,

$$|\nabla_{x_2} p(s, \boldsymbol{x}; t, \boldsymbol{y})| \leqslant C_3 (t-s)^{-\frac{3}{2}} g_{\lambda_3}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big). \tag{1.24}$$

(iv) (Hölder estimate of $\nabla_{x_2} p$ in $\boldsymbol{x}$) For any $\eta_3 \in (0, (\alpha - \gamma) \wedge (\alpha + \gamma - 1))$, there exist constants $\lambda_4,\ C_4 \geqslant 1$ depending on Θ_T, α and η_3 such that for any $0 \leqslant s < t \leqslant T$ and $\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{y} \in \mathbb{R}^{2d}$,

$$\begin{aligned} \big|\nabla_{x_2} p(s, \boldsymbol{x}; t, \boldsymbol{y}) - \nabla_{x_2} p(s, \boldsymbol{x}'; t, \boldsymbol{y})\big| &\leqslant C_4 |\boldsymbol{x} - \boldsymbol{x}'|^{\eta_3}_{\mathbf{d}} (t-s)^{-\frac{3+\eta_3}{2}} \\ &\times \Big(g_{\lambda_4}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big) + g_{\lambda_4}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}') - \boldsymbol{y}\big) \Big). \end{aligned} \tag{1.25}$$

Remark 1.5. For gradient estimates in the degenerate component x_2, we need extra regularities (1.22) and (1.23) since for kinetic operators, we only have $\frac{2}{3}$-gain of regularity in x_2. Let us briefly comment this additional regularity, which might not seem sharp at first sight. Our feeling is that such assumptions are actually sharp with respect to the methodology employed here. Indeed, as our starting point to estimate the density relies on a first order parametrix expansion, see (3.16), we end up with an implicit representation of the density involving as well all the coefficients of the system. The crucial point is that, when estimating the gradient in the degenerate directions of the implicit representation of the density, we make all the coefficients feel the differentiation w.r.t. the degenerate variables. This roughly explains why we impose assumptions (1.22), (1.23) which lead to similar (w.r.t. the degenerate variables) conditions as the one previously imposed on F_2 in $(\mathbf{H}^{\gamma}_{\mathbf{F}})$. Assuming the same regularity in the degenerate directions for the whole drift $\mathbf{F} = (F_1, F_2)$ already appeared in [4] in connection with weak uniqueness for (1.1) as well as in [2] and [14] to derive Schauder like estimates for strong uniqueness purposes. Note that in those frameworks, since the diffusion coefficient was assumed to be Lipschitz, assumption (1.22) did not explicitly appears. Let us eventually conclude by emphasizing that, under similar assumptions as the one of Theorem 1.1, the Authors in [5] only succeeded in deriving gradient estimates in the non degenerate directions, but only Hölder estimates in the degenerate ones.

For two quantities Q_1 and Q_2, we will frequently use the notation $Q_1 \lesssim Q_2$ meaning that that there exists $C := C(\Theta_T)$ such that $Q_1 \leqslant CQ_2$.

2 Preliminaries

In this section we assume that $\mathbf{F} = (F_1, F_2)$ satisfies $(\mathbf{H}^0_{\mathbf{F}})$ and temporarily assume that

$$\|\nabla_{\boldsymbol{x}}\mathbf{F}\|_\infty < \infty.$$

In particular, for some $\kappa_1 > 0$,

$$|F_1(t, \boldsymbol{x}) - F_1(t, \boldsymbol{y})| \leqslant \kappa_1(1 + |\boldsymbol{x} - \boldsymbol{y}|), \tag{2.1}$$

and for some $\kappa_2 > 0$,

$$|F_2(t, \boldsymbol{x}) - F_2(t, \boldsymbol{y})| \leqslant \kappa_2(|(\boldsymbol{x} - \boldsymbol{y})_1| + |(\boldsymbol{x} - \boldsymbol{y})_2|^{\frac{1}{3}} + |(\boldsymbol{x} - \boldsymbol{y})_2|). \tag{2.2}$$

For $s, t \geqslant 0$ and $\boldsymbol{x} \in \mathbb{R}^{2d}$, let $\boldsymbol{\theta}_{t,s}(\boldsymbol{x})$ be the regularization flow defined by the differential system

$$\dot{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) = \mathbf{F}(t, \boldsymbol{\theta}_{t,s}(\boldsymbol{x})), \quad \boldsymbol{\theta}_{s,s}(\boldsymbol{x}) = \boldsymbol{x}. \tag{2.3}$$

Here $(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}))_{t\geqslant s}$ stands for a forward flow, while $(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}))_{t\leqslant s}$ stands for a backward flow. In particular, let $(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}))^{-1}$ be the inverse of $x \mapsto \boldsymbol{\theta}_{t,s}(\boldsymbol{x})$. Then

$$(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}))^{-1} = \boldsymbol{\theta}_{s,t}(\boldsymbol{x}). \tag{2.4}$$

2.1 Equivalence of measurable flow for ODEs

We recall the following (sublinear) Gronwall type lemma.

Lemma 2.1. *Let $f(t) : [0, \infty) \to [0, \infty)$ be a continuous function. Suppose that for some $\alpha \in (0, 1)$ and $c_1, c_2 \geqslant 0$,*

$$f(t) \leqslant f(0) + c_1 \int_0^t f(s)^\alpha \mathrm{d}s + c_2 \int_0^t f(s)\mathrm{d}s,\ \ t > 0.$$

Then

$$f(t) \leqslant \mathrm{e}^{c_2 t} f(0) + (c_1 \mathrm{e}^{c_2 t}(1 - \alpha)t)^{\frac{1}{1-\alpha}},\ \ t > 0.$$

We have the following crucial lemma, which corresponds to Lemma 1.1 of [31] and [6].

Lemma 2.2. *Under $(\mathbf{H}^0_{\mathbf{F}})$, for any $T > 0$, there exist a constant $\kappa_3 \geqslant 1$ only depending on κ_1, κ_2, d, T such that for all $0 \leqslant s \leqslant r < t \leqslant T$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2d}$,*

$$\kappa_3^{-1}\big(|\mathbb{T}^{-1}_{t-s}(\boldsymbol{x} - \boldsymbol{\theta}_{r,t}(\boldsymbol{y}))| - 1\big) \leqslant |\mathbb{T}^{-1}_{t-s}(\boldsymbol{\theta}_{t,r}(\boldsymbol{x}) - \boldsymbol{y})| \leqslant \kappa_3\big(|\mathbb{T}^{-1}_{t-s}(\boldsymbol{x} - \boldsymbol{\theta}_{r,t}(\boldsymbol{y}))| + 1\big). \tag{2.5}$$

Proof. To show (2.5), by (2.4) and the symmetry, it suffices to show

$$|\mathbb{T}^{-1}_{t-s}(\boldsymbol{\theta}_{t,r}(\boldsymbol{x}) - \boldsymbol{\theta}_{t,r}(\boldsymbol{y}))| \leqslant \kappa_3\big(|\mathbb{T}^{-1}_{t-s}(\boldsymbol{x} - \boldsymbol{y})| + 1\big).$$

Without loss of generality, we may assume $0 = r < t \leqslant T$, and write for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2d}$ and $t \geqslant 0$,

$$\ell_i(t) := \big|\big(\boldsymbol{\theta}_{t,0}(\boldsymbol{x}) - \boldsymbol{\theta}_{t,0}(\boldsymbol{y})\big)_i\big|, \quad i = 1, 2.$$

Using the above notation and by definition (1.9), we only need to show that for all $0 < t \leqslant t_0 \leqslant T$,

$$t_0\ell_1(t) + \ell_2(t) \leqslant \kappa_3\big(t_0\ell_1(0) + \ell_2(0) + t_0^{\frac{3}{2}}\big). \tag{2.6}$$

For $i = 1$, by (2.1) we have

$$\begin{aligned}\ell_1(t) &\leqslant \ell_1(0) + \int_0^t |F_1(r, \boldsymbol{\theta}_{r,0}(\boldsymbol{x})) - F_1(r, \boldsymbol{\theta}_{r,0}(\boldsymbol{y}))|\mathrm{d}r \\ &\leqslant \ell_1(0) + \kappa_1 t + \kappa_1 \int_0^t (\ell_1(r) + \ell_2(r))\mathrm{d}r,\end{aligned}$$

which implies by Gronwall's inequality that

$$\ell_1(t) \leqslant \mathrm{e}^{\kappa_1 t}(\ell_1(0) + \kappa_1 t) + \kappa_1 \mathrm{e}^{\kappa_1 t} \int_0^t \ell_2(r)\mathrm{d}r. \tag{2.7}$$

For $i = 2$, by (2.2) and (2.7), we have

$$\begin{aligned}\ell_2(t) &\leqslant \ell_2(0) + \int_0^t |F_2(r, \boldsymbol{\theta}_{r,0}(\boldsymbol{x})) - F_2(r, \boldsymbol{\theta}_{r,0}(\boldsymbol{y}))|\mathrm{d}r \\ &\leqslant \ell_2(0) + \int_0^t \ell_1(r)\mathrm{d}r + \kappa_2 \int_0^t \big(\ell_2(r)^{\frac{1}{3}} + \ell_2(r)\big)\mathrm{d}r \\ &\leqslant \ell_2(0) + \mathrm{e}^{\kappa_1 t}(t\ell_1(0) + \kappa_1 t^2) + \kappa_2 \int_0^t \ell_2(r)^{\frac{1}{3}}\mathrm{d}r + (\kappa_2 + e^{\kappa_1 t}) \int_0^t \ell_2(r)\mathrm{d}r.\end{aligned} \tag{2.8}$$

In particular, by Lemma 2.1, for $\tilde{\kappa}_2 = \kappa_2 + e^{\kappa_1 t}$, we have

$$\begin{aligned}\ell_2(t) &\leqslant \mathrm{e}^{\tilde{\kappa}_2 t}\big(\ell_2(0) + \mathrm{e}^{\kappa_1 t}(t\ell_1(0) + t^2)\big) + (\kappa_2 \mathrm{e}^{\tilde{\kappa}_2 t}\tfrac{2}{3}t)^{\frac{3}{2}} \\ &\leqslant \mathrm{e}^{\tilde{\kappa}_2 t}\ell_2(0) + \mathrm{e}^{(\tilde{\kappa}_2+\kappa_1)t}t\ell_1(0) + c_2 t^{\frac{3}{2}},\end{aligned}$$

which together with (2.7) yields (2.6). □

Remark 2.3. By (2.5) and the flow property $\boldsymbol{\theta}_{t,s}(\boldsymbol{x}) = \boldsymbol{\theta}_{t,r} \circ \boldsymbol{\theta}_{r,s}(\boldsymbol{x})$, we also have

$$|\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{r,s}(\boldsymbol{x}) - \boldsymbol{\theta}_{r,t}(\boldsymbol{y}))| - 1 \lesssim |\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}) - \boldsymbol{y})| \lesssim |\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{r,s}(\boldsymbol{x}) - \boldsymbol{\theta}_{r,t}(\boldsymbol{y}))| + 1. \tag{2.9}$$

Moreover, if $|\boldsymbol{x} - \boldsymbol{x}'|_{\mathbf{d}} \leqslant C_0(t-s)^{1/2}$ for some $C_0 > 0$, then

$$|\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}') - \boldsymbol{y})| - 1 \lesssim |\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}) - \boldsymbol{y})| \lesssim |\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}') - \boldsymbol{y})| + 1. \tag{2.10}$$

Indeed, by (2.5) we have

$$\begin{aligned}|\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}') - \boldsymbol{y})| &\lesssim |\mathbb{T}_{t-s}^{-1}(\boldsymbol{x}' - \boldsymbol{\theta}_{s,t}(\boldsymbol{y}))| + 1 \\ &\leqslant |\mathbb{T}_{t-s}^{-1}(\boldsymbol{x} - \boldsymbol{\theta}_{s,t}(\boldsymbol{y}))| + |\mathbb{T}_{t-s}^{-1}(\boldsymbol{x}' - \boldsymbol{x})| + 1 \\ &\leqslant |\mathbb{T}_{t-s}^{-1}(\boldsymbol{x} - \boldsymbol{\theta}_{s,t}(\boldsymbol{y}))| + C_0 + 1 \\ &\lesssim |\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}) - \boldsymbol{y})| + 1.\end{aligned}$$

The other inequality in (2.9) is derived by symmetry.

2.2 Gram matrix

Let $A_t, \sigma_t : \mathbb{R}_+ \to \mathbb{R}^d \times \mathbb{R}^d$ be two measurable maps. Suppose that for some closed convex subset $\mathcal{E}$ of $GL(\mathbb{R}^d)$ and $\kappa_0 > 0$, and for all $t \geqslant 0$,

$$A_t \in \mathcal{E}, \quad \kappa_0^{-1}|\xi| \leqslant |\sigma_t \xi| \leqslant \kappa_0 |\xi|. \tag{2.11}$$

Define

$$\mathbf{A}_t := \begin{pmatrix} 0_{d\times d} & 0_{d\times d} \\ A_t & 0_{d\times d} \end{pmatrix}, \quad \mathbf{\Sigma}_t := \begin{pmatrix} \sigma_t \\ 0_{d\times d} \end{pmatrix}.$$

For $s, t \geqslant 0$, let $\mathbf{R}_{t,s}$ be the resolvent of $\mathbf{A}_t$, that is,

$$\partial_t \mathbf{R}_{t,s} = \mathbf{A}_t \mathbf{R}_{t,s}, \quad \mathbf{R}_{s,s} = \mathbb{I}_{2d\times 2d}. \tag{2.12}$$

It is easy to see that the unique solution of (2.12) is given by

$$\mathbf{R}_{t,s} = \begin{pmatrix} \mathbb{I}_{d\times d} & 0_{d\times d} \\ \int_s^t A_r \mathrm{d}r & \mathbb{I}_{d\times d} \end{pmatrix}. \tag{2.13}$$

From this expression, one sees that for any $s, t, r \geqslant 0$,

$$\mathbf{R}_{t,s}^{-1} = \mathbf{R}_{s,t}, \quad \mathbf{R}_{t,r}\mathbf{R}_{r,s} = \mathbf{R}_{t,s}. \tag{2.14}$$

The Gram matrix associated with $\mathbf{A}_t$ and $\mathbf{\Sigma}_t$ is defined by

$$\mathbf{K}_{t,s} := \int_s^t \mathbf{R}_{t,r}\mathbf{\Sigma}_r\mathbf{\Sigma}_r^*\mathbf{R}_{t,r}^* \mathrm{d}r.$$

The following lemma is well-known. For reader's convenience, we provide detailed proofs here.

Lemma 2.4. *Under (2.11), there is a constant $\kappa \geqslant 1$ depending only on κ_0 and $\mathcal{E}$ such that for all $0 \leqslant s < t < \infty$ and $\boldsymbol{x} \in \mathbb{R}^{2d}$,*

$$|\mathbf{K}_{t,s}^{-1/2}\boldsymbol{x}|^2 = \langle \mathbf{K}_{t,s}^{-1}\boldsymbol{x}, \boldsymbol{x}\rangle \asymp_\kappa |\mathbb{T}_{t-s}^{-1}\boldsymbol{x}|^2, \tag{2.15}$$

and

$$|\mathbb{T}_{t-s}^{-1}\mathbf{R}_{t,s}\boldsymbol{x}| \asymp_\kappa |\mathbb{T}_{t-s}^{-1}\boldsymbol{x}|. \tag{2.16}$$

Proof. By the definition and the change of variables, it is easy to see that

$$\mathbf{K}_{t,s} = \mathbb{T}_{t-s}\hat{\mathbf{K}}_{1,0}\mathbb{T}_{t-s}, \tag{2.17}$$

where

$$\hat{\mathbf{K}}_{1,0} := \int_0^1 \hat{\mathbf{R}}_{1,r}\hat{\mathbf{\Sigma}}_r\hat{\mathbf{\Sigma}}_r^*\hat{\mathbf{R}}_{1,r}^* \mathrm{d}r,$$

and

$$\hat{\mathbf{\Sigma}}_r = \mathbf{\Sigma}_{s+(t-s)r}, \quad \hat{\mathbf{R}}_{1,r} = \begin{pmatrix} \mathbb{I}_{d\times d} & 0_{d\times d} \\ \int_r^1 A_{s+(t-s)u}\mathrm{d}u & \mathbb{I}_{d\times d} \end{pmatrix}.$$

Thus, without loss of generality, we may assume $s=0$, $t=1$ and $|\boldsymbol{x}|^2=|x_1|^2+|x_2|^2=1$. Clearly,

$$\langle \mathbf{K}_{1,0}^{-1}\boldsymbol{x},\boldsymbol{x}\rangle \asymp 1 \Leftrightarrow \langle \mathbf{K}_{1,0}\boldsymbol{x},\boldsymbol{x}\rangle \asymp 1. \tag{2.18}$$

Note that by (2.11),

$$\langle \mathbf{K}_{1,0}\boldsymbol{x},\boldsymbol{x}\rangle = \int_0^1 |\boldsymbol{\Sigma}_r^*\hat{\mathbf{R}}_{1,r}^*\boldsymbol{x}|^2 \mathrm{d}r = \int_0^1 \left|\sigma_r^*\left[x_1+\int_r^1 A_u^* x_2 \mathrm{d}u\right]\right|^2 \mathrm{d}r \asymp \int_0^1 \left|x_1+\int_r^1 A_u^* x_2 \mathrm{d}u\right|^2 \mathrm{d}r,$$

and by the change of variable,

$$\int_0^1 \left|x_1+\int_r^1 A_u^* x_2 \mathrm{d}u\right|^2 \mathrm{d}r = \int_0^1 \left|x_1+\int_0^r A_{1-u}^* x_2 \mathrm{d}u\right|^2 \mathrm{d}r.$$

Since $A_u\in\mathcal{E}$ and $\mathcal{E}$ is a closed convex subset of $GL(\mathbb{R}^d)$, we have for some $c_0\in(0,1)$,

$$\inf_{A\in\mathcal{E}} |A^* x_2| \geqslant c_0|x_2| \Rightarrow c_0 r|x_2| \leqslant \left|\int_0^r A_u^* x_2 \mathrm{d}u\right| \leqslant c_0^{-1} r|x_2|.$$

Recall $|x_1|^2+|x_2|^2=1$. If $|x_1|\leqslant \frac{c_0^3}{4}|x_2|$, then $|x_2|^2\geqslant (\frac{c_0^6}{16}+1)^{-1}$ and

$$\begin{aligned}\int_0^1 \left|x_1+\int_0^r A_u^* x_2 \mathrm{d}u\right|^2 \mathrm{d}r &\geqslant \int_0^1 \left(|x_1|^2-2c_0^{-1}r|x_1||x_2|+c_0^2r^2|x_2|^2\right)\mathrm{d}r\\ &\geqslant c_0^2|x_2|^2\left(\int_0^1 \left[r^2-\frac{r}{2}\right]\mathrm{d}r\right) \geqslant \frac{c_0^2}{12}\Big(\frac{c_0^6}{16}+1\Big)^{-1};\end{aligned}$$

if $|x_1|\geqslant \frac{c_0^3}{4}|x_2|$, then $|x_1|^2\geqslant (1+\frac{16}{c_0^6})^{-1}$ and

$$\int_0^1 \left|x_1+\int_0^r A_u^* x_2 \mathrm{d}u\right|^2 \mathrm{d}r \geqslant \int_0^{c_0^4/8} \big||x_1|-rc_0^{-1}|x_2|\big|^2 \mathrm{d}r \geqslant |x_1|^2\int_0^{c_0^4/8}(1-r4c_0^{-4})^2\mathrm{d}r \geqslant \frac{c_0^4}{32}\Big(1+\frac{16}{c_0^6}\Big)^{-1}.$$

We thus obtain (2.18). As for (2.16), it then readily follows from the scaling relation (2.17). □

2.3 Control problem

In this subsection we show how the quantity $|\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x})-\boldsymbol{y})|^2$ appearing in our main estimates can be, under (2.1) and (2.2), related to a control problem (see [9]) associated with $\mathbf{F}$ and B. More precisely, we consider the following deterministic control problem:

$$\dot{\boldsymbol{\phi}}_{r,s} = \mathbf{F}(r,\boldsymbol{\phi}_{r,s})+B\varphi_r,\quad r\in[s,t],\quad \boldsymbol{\phi}_{s,s}=\boldsymbol{x},\quad \boldsymbol{\phi}_{t,s}=\boldsymbol{y}, \tag{2.19}$$

where $\varphi:[s,t]\to\mathbb{R}^d$ is a square integrable control function. Let $I(s,\boldsymbol{x};t,\boldsymbol{y})$ be the associated energy functional

$$I(s,\boldsymbol{x};t,\boldsymbol{y}) = \inf\left\{\left(\int_s^t |\varphi_r|^2\mathrm{d}r\right)^{1/2},\quad \boldsymbol{\phi}_{s,s}=\boldsymbol{x},\quad \boldsymbol{\phi}_{t,s}=\boldsymbol{y}\right\},$$

where the infimum is taken over all admissible controls φ. The following proposition plays a crucial role for proving the lower bound estimate of the heat kernel.

Proposition 2.5. *Under* $(\mathbf{H}^0_{\mathbf{F}})$*, for any* $T>0$*, there exist constants* $\kappa_5,\kappa_6\geqslant 1$ *depending only on* $T,\kappa_0,\kappa_1,d,\mathcal{E}$ *such that for all* $0\leqslant s<t\leqslant T$ *and* $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^{2d}$*,*

$$\kappa_5^{-1}\big(|\mathbb{T}^{-1}_{t-s}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x})-\boldsymbol{y})|-1\big)\leqslant I(s,\boldsymbol{x};t,\boldsymbol{y})\leqslant\kappa_5\big(|\mathbb{T}^{-1}_{t-s}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x})-\boldsymbol{y})|+1\big).\tag{2.20}$$

Moreover, one can find a control $\varphi:[s,t]\to\mathbb{R}^d$ *and a solution* $\phi_{r,s}$ *to ODE* (2.19) *such that*

$$\sup_{r\in[s,t]}|\varphi_r|\leqslant\kappa_6\big(|\mathbb{T}^{-1}_{t-s}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x})-\boldsymbol{y})|+1\big)/\sqrt{t-s}.\tag{2.21}$$

Proof. Without loss of generality, we may assume $s=0$ and $t=t_0\leqslant\delta$, where δ is a small number only depending on $T,\kappa_0,\kappa_1,d,\mathcal{E}$. Let $(\boldsymbol{\phi}_t)_{t\in[0,t_0]}$ be any solution of control problem (2.19). Let

$$\boldsymbol{\psi}_t:=\boldsymbol{\phi}_t-\boldsymbol{\theta}_t,\quad \boldsymbol{\theta}_t:=\boldsymbol{\theta}_{t,0}(\boldsymbol{x}).$$

Then $\boldsymbol{\psi}_t=(\psi^1_t,\psi^2_t)$ solves the following control problem:

$$\dot{\boldsymbol{\psi}}_t=\mathbf{A}(t,\boldsymbol{\psi}_t)\psi_t+\widetilde{\mathbf{F}}(t,\boldsymbol{\psi}_t)+B\varphi_t,\quad\boldsymbol{\psi}_0=0,\quad\boldsymbol{\psi}_{t_0}=\boldsymbol{y}-\boldsymbol{\theta}_{t_0},\tag{2.22}$$

where

$$\widetilde{\mathbf{F}}(s,\boldsymbol{x}):=\begin{pmatrix}F_1(s,\boldsymbol{x}+\boldsymbol{\theta}_t)-F_1(t,\boldsymbol{\theta}_t)\\ F_2(t,\boldsymbol{\theta}^1_t,x_2+\boldsymbol{\theta}^2_t)-F_2(t,\boldsymbol{\theta}^1_t,\boldsymbol{\theta}^2_t)\end{pmatrix},$$

and

$$\mathbf{A}(s,\boldsymbol{x}):=\begin{pmatrix}0_{d\times d}&0_{d\times d}\\ \int_0^1\nabla_{x_1}F_2(t,ux_1+\boldsymbol{\theta}^1_t,x_2+\boldsymbol{\theta}^2_t)\mathrm{d}u&0_{d\times d}\end{pmatrix}.$$

By (2.22) and (2.1), we have

$$|\psi^1_t|\lesssim\int_0^t(1+|\psi_s|)\mathrm{d}s+\int_0^t|\varphi_s|\mathrm{d}s,$$

and due to $\nabla_{x_1}F_2\in\mathcal{E}$ and by (2.2),

$$|\psi^2_t|\lesssim\int_0^t|\psi^1_s|\mathrm{d}s+\int_0^t(|\psi^2_s|^{\frac13}+|\psi^2_s|)\mathrm{d}s.$$

Thus by Lemma 2.1 we have

$$|\psi^1_t|\lesssim t+\int_0^t|\varphi_s|\mathrm{d}s,\quad|\psi^2_t|\lesssim t^{3/2}+t\int_0^t|\varphi_s|\mathrm{d}s,\ t\in[0,t_0].\tag{2.23}$$

Hence,

$$|\mathbb{T}^{-1}_{t_0}(\boldsymbol{y}-\boldsymbol{\theta}_{t_0})|=t_0^{-\frac12}|\psi^1_{t_0}|+t_0^{-\frac32}|\psi^2_{t_0}|\lesssim 1+t_0^{-\frac12}\int_0^{t_0}|\varphi_s|\mathrm{d}s\leqslant 1+\left(\int_0^{t_0}|\varphi_s|^2\mathrm{d}s\right)^{1/2},$$

which gives the left hand side estimate in (2.20).

On the other hand, by Coron [Theorem 3.40], system (2.22) is controllable, and the exhibited control φ is given by

$$\varphi_s=(\mathbf{R}_{t_0,s}B)^*\mathbf{K}^{-1}_{t_0,0}\left(\boldsymbol{y}-\boldsymbol{\theta}_{t_0}-\int_0^{t_0}\mathbf{R}_{t_0,s}\widetilde{\mathbf{F}}(s,\psi_s)\mathrm{d}s\right),\tag{2.24}$$

where $\mathbf{R}_{t,s}$ is the resolvent of $\mathbf{A}(t,\psi_t)$ (see (2.12)), and

$$\mathbf{K}_{t_0,0} := \int_0^{t_0} \mathbf{R}_{t_0,s} BB^* \mathbf{R}^*_{t_0,s} \mathrm{d}s.$$

Indeed, by Duhamel's formula, the above control φ satisfies (2.22). Note that by (2.13),

$$\mathbf{R}_{t_0,s}B = \begin{pmatrix} \mathbb{I}_{d\times d} \\ \int_s^{t_0}\int_0^1 \nabla_{x_1} F_2(r, u\psi^1_r + \boldsymbol{\theta}^1_r, \psi^2_r + \boldsymbol{\theta}^2_r)\mathrm{d}u\mathrm{d}r \end{pmatrix}.$$

By (2.17) and (2.15), it is easy to see that

$$\begin{aligned}
\sup_{s\in[0,t_0]} |\varphi_s| &\lesssim t_0^{-\frac{1}{2}} \left(|\mathbb{T}^{-1}_{t_0}(\boldsymbol{y} - \boldsymbol{\theta}_{t_0})| + \left| \mathbb{T}^{-1}_{t_0} \int_0^{t_0} \mathbf{R}_{t_0,s} \widetilde{\mathbf{F}}(s,\psi_s)\mathrm{d}s \right| \right) \\
&\overset{(2.2)}{\lesssim} t_0^{-\frac{1}{2}} |\mathbb{T}^{-1}_{t_0}(\boldsymbol{y} - \boldsymbol{\theta}_{t_0})| + t_0^{-1} \int_0^{t_0} (1 + |\psi_s|)\mathrm{d}s + t_0^{-2} \int_0^{t_0} (|\psi^2_s|^{\frac{1}{3}} + |\psi^2_s|)\mathrm{d}s \\
&\overset{(2.23)}{\lesssim} t_0^{-\frac{1}{2}} |\mathbb{T}^{-1}_{t_0}(\boldsymbol{y} - \boldsymbol{\theta}_{t_0})| + t_0^{-\frac{1}{2}} + t_0 \sup_{s\in[0,t_0]} |\varphi_s| + t_0^{-\frac{1}{3}} \sup_{s\in[0,t_0]} |\varphi_s|^{\frac{1}{3}} \\
&\lesssim t_0^{-\frac{1}{2}} |\mathbb{T}^{-1}_{t_0}(\boldsymbol{y} - \boldsymbol{\theta}_{t_0})| + \varepsilon^{-1} t_0^{-\frac{1}{2}} + (t_0 + \varepsilon) \sup_{s\in[0,t_0]} |\varphi_s|,
\end{aligned}$$

where the last step is due to Young's inequality and the implicit constant only depends on $T, \kappa_0, \kappa_1, d, \mathcal{E}$. In particular, we can choose ε and δ small enough so that for all $t_0 \in (0,\delta]$,

$$\sup_{s\in[0,t_0]} |\varphi_s| \lesssim t_0^{-\frac{1}{2}} \big(|\mathbb{T}^{-1}_{t_0}(\boldsymbol{y} - \boldsymbol{\theta}_{t_0})| + 1 \big).$$

This in turn yields (2.21) as well as the right hand side estimate in (2.20). $\square$

Remark 2.6. When $\mathbf{F}$ is Lipschitz continuous, [9] has shown that

$$I(s,\boldsymbol{x};t,\boldsymbol{y}) \asymp |\mathbb{T}^{-1}_{t-s}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}) - \boldsymbol{y})|$$

and

$$\sup_{r\in[s,t]} |\varphi_r| \lesssim_C |\mathbb{T}^{-1}_{t-s}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}) - \boldsymbol{y})| / \sqrt{t-s}.$$

The additional constant in the estimates of Proposition 2.5, due to the rougher framework, will anyhow not perturb too much the analysis.

3 Density bounds for SDEs with smooth coefficients

In this section we always assume $(\mathbf{H}^\gamma_\sigma)$ and $(\mathbf{H}^\gamma_{\mathbf{F}})$ for some $\gamma \in (0,1)$, and temporarily assume that

$$\|\nabla^j_{\boldsymbol{x}} \mathbf{F}\|_\infty < \infty, \quad \|\nabla^j_{\boldsymbol{x}} \sigma\|_\infty < \infty, \quad j \in \mathbb{N}. \tag{3.1}$$

It is well known that in the current smooth coefficients framework there exists a transition density $p(s,\boldsymbol{x};t,\boldsymbol{y})$ which is C_b^∞- smooth in variables $\boldsymbol{x},\boldsymbol{y}$ for all $s<t$, by Hörmander's theorem. Moreover, $p(s,\boldsymbol{x};t,\boldsymbol{y})$ satisfies the following backward Kolmogorov equation

$$\partial_s p(s,\boldsymbol{x};t,\boldsymbol{y})+\mathcal{L}_{s,\boldsymbol{x}}p(s,\boldsymbol{x};t,\boldsymbol{y})=0,\quad p(s,\cdot;t,\boldsymbol{y})\longrightarrow\delta_{\boldsymbol{y}}(\cdot)\text{ weakly as }s\uparrow t, \tag{3.2}$$

and the forward Kolmogorov equation (Fokker-Planck equation):

$$\partial_t p(s,\boldsymbol{x};t,\boldsymbol{y})-\mathcal{L}^*_{(t,\boldsymbol{y})}p(s,\boldsymbol{x};t,\boldsymbol{y})=0,\quad p(s,\boldsymbol{x};t,\cdot)\longrightarrow\delta_{\boldsymbol{x}}(\cdot)\text{ weakly as }t\downarrow s, \tag{3.3}$$

where, setting $a=\sigma\sigma^*/2$,

$$\mathcal{L}_{s,\boldsymbol{x}}f(\boldsymbol{x})=\operatorname{tr}\big(a(s,\boldsymbol{x})\nabla^2_{x_1}f(\boldsymbol{x})\big)+\langle\mathbf{F}(s,\boldsymbol{x}),\nabla_{\boldsymbol{x}}f(\boldsymbol{x})\rangle,$$

and

$$\mathcal{L}^*_{(t,\boldsymbol{y})}f(\boldsymbol{y})=\operatorname{tr}\big(\nabla^2_{y_1}(a(t,\boldsymbol{y})f(\boldsymbol{y}))\big)-\operatorname{div}_{\boldsymbol{y}}(\mathbf{F}(t,\boldsymbol{y})f(\boldsymbol{y})).$$

The scope of the section is to obtain two-sided Aronson like bounds, where all the constants appearing below only depend on Θ_T.

3.1 The Duhamel representation for $p(s,\boldsymbol{x};t,\boldsymbol{y})$

Fix now $(\tau,\boldsymbol{\xi})\in\mathbb{R}_+\times\mathbb{R}^{2d}$ as *freezing parameters* to be chosen later on and let

$$\dot{\boldsymbol{\theta}}_{t,\tau}(\boldsymbol{\xi})=\mathbf{F}(t,\boldsymbol{\theta}_{t,\tau}(\boldsymbol{\xi})),\quad t\geqslant 0,\quad \boldsymbol{\theta}_{\tau,\tau}(\boldsymbol{\xi})=\boldsymbol{\xi}. \tag{3.4}$$

We consider the stochastic linearized dynamics $(\widetilde{\mathbf{X}}^{(\tau,\boldsymbol{\xi})}_{t,s})_{t\geqslant s}$:

$$\widetilde{\mathbf{X}}^{(\tau,\boldsymbol{\xi})}_{t,s}=\boldsymbol{x}+\int_s^t[\mathbf{F}(r,\boldsymbol{\theta}_{r,\tau}(\boldsymbol{\xi}))+\mathbf{A}(r,\boldsymbol{\theta}_{r,\tau}(\boldsymbol{\xi}))(\widetilde{\mathbf{X}}^{(\tau,\boldsymbol{\xi})}_{r,s}-\boldsymbol{\theta}_{r,\tau}(\boldsymbol{\xi}))]\mathrm{d}r+\int_s^t B\sigma(r,\boldsymbol{\theta}_{r,\tau}(\boldsymbol{\xi}))\mathrm{d}W_r, \tag{3.5}$$

where, for all $\boldsymbol{x}=(x_1,x_2)\in\mathbb{R}^{2d}$,

$$\mathbf{A}(r,\boldsymbol{x})=\begin{pmatrix}0_{d\times d} & 0_{d\times d}\\ \nabla_{x_1}F_2(r,\boldsymbol{x}) & 0_{d\times d}\end{pmatrix}.$$

Let $(\mathbf{R}^{(\tau,\boldsymbol{\xi})}_{t,s})_{t\geqslant s}$ be the resolvent associated with $\mathbf{A}(r,\boldsymbol{\theta}_{r,\tau}(\boldsymbol{\xi}))$ (see (2.12)), which is explicitly given by

$$\mathbf{R}^{(\tau,\boldsymbol{\xi})}_{t,s}=\begin{pmatrix}\mathbb{I}_{d\times d} & 0_{d\times d}\\ \int_s^t\nabla_{x_1}F_2(r,\boldsymbol{\theta}_{r,\tau}(\boldsymbol{\xi}))\mathrm{d}r & \mathbb{I}_{d\times d}\end{pmatrix}. \tag{3.6}$$

If we define

$$\boldsymbol{\vartheta}^{(\tau,\boldsymbol{\xi})}_{t,s}(\boldsymbol{x}):=\mathbf{R}^{(\tau,\boldsymbol{\xi})}_{t,s}\boldsymbol{x}+\int_s^t\mathbf{R}^{(\tau,\boldsymbol{\xi})}_{t,r}\Big(\mathbf{F}(r,\boldsymbol{\theta}_{r,\tau}(\boldsymbol{\xi}))-\mathbf{A}(r,\boldsymbol{\theta}_{r,\tau}(\boldsymbol{\xi}))\boldsymbol{\theta}_{r,\tau}(\boldsymbol{\xi})\Big)\mathrm{d}r \tag{3.7}$$

and

$$\boldsymbol{\Sigma}^{(\tau,\boldsymbol{\xi})}_r:=B\sigma(r,\boldsymbol{\theta}_{r,\tau}(\boldsymbol{\xi})), \tag{3.8}$$

then by the variation formula of constants, $\widetilde{\mathbf{X}}^{(\tau,\boldsymbol{\xi})}_{t,s}(\boldsymbol{x})$ is explicitly given by

$$\widetilde{\mathbf{X}}^{(\tau,\boldsymbol{\xi})}_{t,s}(\boldsymbol{x}) = \boldsymbol{\vartheta}^{(\tau,\boldsymbol{\xi})}_{t,s}(\boldsymbol{x}) + \int_s^t \mathbf{R}^{(\tau,\boldsymbol{\xi})}_{t,r}\boldsymbol{\Sigma}^{(\tau,\boldsymbol{\xi})}_r \mathrm{d}W_r. \tag{3.9}$$

Clearly, the random variable $\widetilde{\mathbf{X}}^{(\tau,\boldsymbol{\xi})}_{t,s}(\boldsymbol{x})$ admits a Gaussian density $\widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};t,\cdot)$ given by

$$\widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};t,\boldsymbol{y}) = \frac{1}{(2\pi)^d \det(\mathbf{K}^{(\tau,\boldsymbol{\xi})}_{t,s})^{\frac12}} \exp\left(-\frac12\left|(\mathbf{K}^{(\tau,\boldsymbol{\xi})}_{t,s})^{-\frac12}(\boldsymbol{\vartheta}^{(\tau,\boldsymbol{\xi})}_{t,s}(\boldsymbol{x})-\boldsymbol{y})\right|^2\right), \tag{3.10}$$

where

$$\mathbf{K}^{(\tau,\boldsymbol{\xi})}_{t,s} := \int_s^t \mathbf{R}^{(\tau,\boldsymbol{\xi})}_{t,r}\boldsymbol{\Sigma}^{(\tau,\boldsymbol{\xi})}_r(\mathbf{R}^{(\tau,\boldsymbol{\xi})}_{t,r}\boldsymbol{\Sigma}^{(\tau,\boldsymbol{\xi})}_r)^*\mathrm{d}r. \tag{3.11}$$

In particular, $\widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};t,\boldsymbol{y})$ satisfies

$$\partial_s\widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};t,\boldsymbol{y}) + \widetilde{\mathscr{L}}^{(\tau,\boldsymbol{\xi})}_{s,\boldsymbol{x}}\widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};t,\boldsymbol{y}) = 0, \quad \widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};t,\boldsymbol{y}) \longrightarrow \delta_{\boldsymbol{y}}(\cdot) \text{ weakly as } s\uparrow t, \tag{3.12}$$

where, for $a=\sigma\sigma^*/2$,

$$\widetilde{\mathscr{L}}^{(\tau,\boldsymbol{\xi})}_{s,\boldsymbol{x}} = \mathrm{tr}\big(a(s,\boldsymbol{\theta}_{s,\tau}(\boldsymbol{\xi}))\cdot\nabla^2_{x_1}\big) + \langle(\mathbf{F}(s,\boldsymbol{\theta}_{s,\tau}(\boldsymbol{\xi})) + \mathbf{A}(s,\boldsymbol{\theta}_{s,\tau}(\boldsymbol{\xi}))(\boldsymbol{x}-\boldsymbol{\theta}_{s,\tau}(\boldsymbol{\xi})))\,,\nabla_{\boldsymbol{x}}\rangle \tag{3.13}$$

denotes the generator of the diffusion with frozen coefficients in (3.5).

The following proposition is a direct consequence of expression (3.10) and Lemma 2.4.

Proposition 3.1 (A priori controls for the frozen Gaussian densities)**.** *Under* $(\mathbf{H}^0_\sigma)$, $(\mathbf{H}^{\mathbf{0}}_{\mathbf{F}})$ *and* (3.1), *for any* $T>0$ *and* $j=(j_1,j_2)\in\mathbb{N}_0^2$, *there are constants* $\lambda_j, C_j \geqslant 1$ *depending only on* Θ_T *such that for all* $0\leqslant s<t\leqslant T$, $\tau\in[0,T]$ *and* $\boldsymbol{x},\boldsymbol{y},\boldsymbol{\xi}\in\mathbb{R}^{2d}$,

$$C_0^{-1}g_{\lambda_0^{-1}}\big(t-s,\boldsymbol{\vartheta}^{(\tau,\boldsymbol{\xi})}_{t,s}(\boldsymbol{x})-\boldsymbol{y}\big) \leqslant \widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};t,\boldsymbol{y}) \leqslant C_0 g_{\lambda_0}\big(t-s,\boldsymbol{\vartheta}^{(\tau,\boldsymbol{\xi})}_{t,s}(\boldsymbol{x})-\boldsymbol{y}\big), \tag{3.14}$$

where $g_\lambda(t,\boldsymbol{x})$ *is defined by* (1.14), *and*

$$|\nabla^{j_1}_{x_1}\nabla^{j_2}_{x_2}\widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};t,\boldsymbol{y})| \leqslant C_j(t-s)^{-\frac{j_1+3j_2}{2}}g_{\lambda_j}\big(t-s,\boldsymbol{\vartheta}^{(\tau,\boldsymbol{\xi})}_{t,s}(\boldsymbol{x})-\boldsymbol{y}\big). \tag{3.15}$$

The starting point of our analysis is the following Duhamel type representation formula which readily follows in the current *smooth coefficients* setting from (3.2)-(3.3) and (3.13):

$$\begin{aligned} p(s,\boldsymbol{x};t,\boldsymbol{y}) &= \widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};t,\boldsymbol{y}) + \int_s^t\int_{\mathbb{R}^{2d}} \widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};r,\boldsymbol{z})(\mathscr{L}_{r,\boldsymbol{z}} - \widetilde{\mathscr{L}}^{(\tau,\boldsymbol{\xi})}_{r,\boldsymbol{z}})p(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r & (3.16)\\ &= \widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};t,\boldsymbol{y}) + \int_s^t\int_{\mathbb{R}^{2d}} p(s,\boldsymbol{x};r,\boldsymbol{z})(\mathscr{L}_{r,\boldsymbol{z}} - \widetilde{\mathscr{L}}^{(\tau,\boldsymbol{\xi})}_{r,\boldsymbol{z}})\widetilde{p}^{(\tau,\boldsymbol{\xi})}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r, & (3.17)\end{aligned}$$

If we take $(\tau,\boldsymbol{\xi})=(s,\boldsymbol{x})$ in (3.16) and set $\widetilde{p}_0(s,\boldsymbol{x};t,\boldsymbol{y}) := \widetilde{p}^{(s,\boldsymbol{x})}(s,\boldsymbol{x};t,\boldsymbol{y})$, then we obtain the *backward* representation

$$p(s,\boldsymbol{x};t,\boldsymbol{y}) = \widetilde{p}_0(s,\boldsymbol{x};t,\boldsymbol{y}) + \int_s^t\int_{\mathbb{R}^{2d}} \widetilde{p}_0(s,\boldsymbol{x};r,\boldsymbol{z})(\mathscr{L}_{r,\boldsymbol{z}} - \widetilde{\mathscr{L}}^{(s,\boldsymbol{x})}_{r,\boldsymbol{z}})p(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r.$$

If we take $(\tau,\boldsymbol{\xi})=(t,\boldsymbol{y})$ in (3.17) and set $\widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y}):=\widetilde{p}^{(t,\boldsymbol{y})}(s,\boldsymbol{x};t,\boldsymbol{y})$, we then obtain the *forward* representation

$$p(s,\boldsymbol{x};t,\boldsymbol{y})=\widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y})+\int_s^t\int_{\mathbb{R}^{2d}}p(s,\boldsymbol{x};r,\boldsymbol{z})(\mathcal{L}_{r,\boldsymbol{z}}-\widetilde{\mathcal{L}}^{(t,\boldsymbol{y})}_{r,\boldsymbol{z}})\widetilde{p}_1(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r.$$

To give the estimates of $\widetilde{p}_i(s,\boldsymbol{x};t,\boldsymbol{y})$, $i=0,1$, we need the following lemma.

Lemma 3.2. *For any $t\geqslant s$ and $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^{2d}$, it holds that*

$$\boldsymbol{\vartheta}^{(s,\boldsymbol{x})}_{t,s}(\boldsymbol{x})=\boldsymbol{\theta}_{t,s}(\boldsymbol{x}),\quad \boldsymbol{\vartheta}^{(t,\boldsymbol{y})}_{t,s}(\boldsymbol{x})-\boldsymbol{y}=\mathbf{R}^{(t,\boldsymbol{y})}_{t,s}(\boldsymbol{x}-\boldsymbol{\theta}_{s,t}(\boldsymbol{y})).$$

Moreover, under $(\mathbf{H}^\gamma_{\mathbf{F}})$, for any $T>0$, there is a constant $C:=C(\Theta_T)>0$ such that for all $0\leqslant s\leqslant t\leqslant T$ and $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^{2d}$,

$$|\mathbb{T}^{-1}_{t-s}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x})-\boldsymbol{\vartheta}^{(t,\boldsymbol{y})}_{t,s}(\boldsymbol{x}))|\leqslant C(t-s)^{\frac{\gamma}{2}}\big(|\mathbb{T}^{-1}_{t-s}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x})-\boldsymbol{y})|^{1+\gamma}+1\big).\tag{3.18}$$

Proof. (i) Since $\boldsymbol{\vartheta}^{(\tau,\boldsymbol{\xi})}_{t,s}(\boldsymbol{x})=\mathbb{E}(\widetilde{\mathbf{X}}^{(\tau,\boldsymbol{\xi})}_{t,s})$, by (3.5) one sees that

$$\boldsymbol{\vartheta}^{(\tau,\boldsymbol{\xi})}_{t,s}(\boldsymbol{x})=\boldsymbol{x}+\int_s^t[\mathbf{F}(r,\boldsymbol{\theta}_{r,\tau}(\boldsymbol{\xi}))+\mathbf{A}(r,\boldsymbol{\theta}_{r,\tau}(\boldsymbol{\xi}))(\boldsymbol{\vartheta}^{(\tau,\boldsymbol{\xi})}_{r,s}(\boldsymbol{x})-\boldsymbol{\theta}_{r,\tau}(\boldsymbol{\xi}))]\mathrm{d}r.\tag{3.19}$$

Hence,

$$\begin{aligned}|\boldsymbol{\vartheta}^{(s,\boldsymbol{x})}_{t,s}(\boldsymbol{x})-\boldsymbol{\theta}_{t,s}(\boldsymbol{x})|&\leqslant\int_s^t|\mathbf{A}(r,\boldsymbol{\theta}_{r,s}(\boldsymbol{x}))(\boldsymbol{\vartheta}^{(s,\boldsymbol{x})}_{r,s}(\boldsymbol{x})-\boldsymbol{\theta}_{r,s}(\boldsymbol{x}))|\mathrm{d}r\\&\leqslant\|\nabla_{x_1}F_2\|_\infty\int_s^t|\boldsymbol{\vartheta}^{(s,\boldsymbol{x})}_{r,s}(\boldsymbol{x})-\boldsymbol{\theta}_{r,s}(\boldsymbol{x})|\mathrm{d}r,\end{aligned}$$

which implies by Gronwall's inequality that

$$\boldsymbol{\vartheta}^{(s,\boldsymbol{x})}_{t,s}(\boldsymbol{x})=\boldsymbol{\theta}_{t,s}(\boldsymbol{x}),$$

which gives the first equality of the lemma.

(ii) For the second one, note that by (3.7) and (2.14),

$$\mathbf{R}^{(t,\boldsymbol{y})}_{s,t}\boldsymbol{\vartheta}^{(t,\boldsymbol{y})}_{t,s}(\boldsymbol{x})=\boldsymbol{x}+\int_s^t\mathbf{R}^{(t,\boldsymbol{y})}_{s,r}\Big(\mathbf{F}(r,\boldsymbol{\theta}_{r,t}(\boldsymbol{y}))-\mathbf{A}(r,\boldsymbol{\theta}_{r,t}(\boldsymbol{y}))\boldsymbol{\theta}_{r,t}(\boldsymbol{y})\Big)\mathrm{d}r.$$

Since $\partial_s\mathbf{R}^{(t,\boldsymbol{y})}_{s,t}=-\mathbf{A}(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{y}))\mathbf{R}^{(t,\boldsymbol{y})}_{s,t}$, by Duhamel's formula, we also have

$$\begin{aligned}\Gamma_s(\boldsymbol{y})&:=\mathbf{R}^{(t,\boldsymbol{y})}_{s,t}\boldsymbol{y}-\int_s^t\mathbf{R}^{(t,\boldsymbol{y})}_{s,r}\Big(\mathbf{F}(r,\boldsymbol{\theta}_{r,t}(\boldsymbol{y}))-\mathbf{A}(r,\boldsymbol{\theta}_{r,t}(\boldsymbol{y}))\boldsymbol{\theta}_{r,t}(\boldsymbol{y})\Big)\mathrm{d}r\\&=\boldsymbol{y}+\int_s^t[\mathbf{F}(r,\boldsymbol{\theta}_{r,t}(\boldsymbol{y}))-\mathbf{A}(r,\boldsymbol{\theta}_{r,t}(\boldsymbol{y}))(\Gamma_r(\boldsymbol{y})-\boldsymbol{\theta}_{r,t}(\boldsymbol{y}))]\mathrm{d}r.\end{aligned}$$

As above, one has $\Gamma_s(\boldsymbol{y})=\boldsymbol{\theta}_{s,t}(\boldsymbol{y})$. Hence,

$$\mathbf{R}^{(t,\boldsymbol{y})}_{s,t}\boldsymbol{\vartheta}^{(t,\boldsymbol{y})}_{t,s}(\boldsymbol{x})-\boldsymbol{x}=\mathbf{R}^{(t,\boldsymbol{y})}_{s,t}\boldsymbol{y}-\boldsymbol{\theta}_{s,t}(\boldsymbol{y}).$$

(iii) Let us now turn to the proof of (3.18). Fix $0 \leqslant s < u \leqslant t \leqslant T$. Note that by (3.19),

$$\boldsymbol{\theta}_{t,s}(\boldsymbol{x}) - \boldsymbol{\vartheta}^{(t,\boldsymbol{y})}_{t,s}(\boldsymbol{x}) = \int_s^t [\mathbf{F}(r,\boldsymbol{\theta}_{r,s}(\boldsymbol{x})) - \mathbf{F}(r,\boldsymbol{\theta}_{r,t}(\boldsymbol{y})) - \mathbf{A}(r,\boldsymbol{\theta}_{r,t}(\boldsymbol{y}))(\boldsymbol{\vartheta}^{(t,\boldsymbol{y})}_{r,s}(\boldsymbol{x}) - \boldsymbol{\theta}_{r,t}(\boldsymbol{y}))]\mathrm{d}r,$$

and by (2.9),

$$\begin{aligned}(t-s)^{-\frac{1}{2}}|(\boldsymbol{\theta}_{r,s}(\boldsymbol{x}) - \boldsymbol{\theta}_{r,t}(\boldsymbol{y}))_1| + (t-s)^{-\frac{3}{2}}|(\boldsymbol{\theta}_{r,s}(\boldsymbol{x}) - \boldsymbol{\theta}_{r,t}(\boldsymbol{y}))_2|\\ \lesssim |\mathbb{T}^{-1}_{t-s}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}) - \boldsymbol{y})| + 1 =: \mathcal{A}.\end{aligned} \tag{3.20}$$

Then we have

$$\begin{aligned}|(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}) - \boldsymbol{\vartheta}^{(t,\boldsymbol{y})}_{t,s}(\boldsymbol{x}))_1| &\leqslant \int_s^t |F_1(r,\boldsymbol{\theta}_{r,s}(\boldsymbol{x})) - F_1(r,\boldsymbol{\theta}_{r,t}(\boldsymbol{y}))|\mathrm{d}r\\ &\lesssim \int_s^t (1 + |\boldsymbol{\theta}_{r,s}(\boldsymbol{x}) - \boldsymbol{\theta}_{r,t}(\boldsymbol{y})|)\mathrm{d}r \lesssim (t-s)\mathcal{A},\end{aligned}$$

and by definition,

$$\begin{aligned}|(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}) - \boldsymbol{\vartheta}^{(t,\boldsymbol{y})}_{t,s}(\boldsymbol{x}))_2| &\leqslant \int_s^t \Big[\|\nabla_{x_1} F_2\|_\infty |(\boldsymbol{\theta}_{r,s}(\boldsymbol{x}) - \boldsymbol{\vartheta}^{(t,\boldsymbol{y})}_{r,s}(\boldsymbol{x}))_1| + |\mathcal{T}_{F_2(r)}(\boldsymbol{\theta}_{r,s}(\boldsymbol{x}), \boldsymbol{\theta}_{r,t}(\boldsymbol{y}))|\Big]\mathrm{d}r\\ &\overset{(1.13)}{\lesssim} \int_s^t \Big[|(\boldsymbol{\theta}_{r,s}(\boldsymbol{x}) - \boldsymbol{\vartheta}^{(t,\boldsymbol{y})}_{r,s}(\boldsymbol{x}))_1| + |\boldsymbol{\theta}_{r,s}(\boldsymbol{x}) - \boldsymbol{\theta}_{r,t}(\boldsymbol{y})|^{1+\gamma}_{\mathbf{d}}\Big]\mathrm{d}r\\ &\overset{(3.20)}{\lesssim} (t-s)^2\mathcal{A} + (t-s)^{\frac{3+\gamma}{2}}\mathcal{A}^{1+\gamma}.\end{aligned}$$

Thus we obtain (3.18). The proof is complete. □

For any $\lambda > 0$, $0 \leqslant s < t < \infty$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2d}$, let

$$\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}) := g_\lambda\big(t-s, \boldsymbol{\theta}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big) = (t-s)^{-2d}\mathrm{e}^{-|\mathbb{T}^{-1}_{t-s}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x})-\boldsymbol{y})|^2/(2\lambda)}, \tag{3.21}$$

with $\boldsymbol{\theta}_{t,s}(\boldsymbol{x})$ the flow associated with $\mathbf{F}$.

Lemma 3.3. *For any $T, \lambda > 0$ and $\alpha > 0$, there are $\lambda' > \lambda$ and $C > 0$ depending only on Θ_T and λ, α such that for all $0 \leqslant s < t \leqslant T$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2d}$,*

$$\big(|\boldsymbol{\theta}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}|^\alpha_{\mathbf{d}} + |\boldsymbol{x} - \boldsymbol{\theta}_{s,t}(\boldsymbol{y})|^\alpha_{\mathbf{d}}\big)\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}) \leqslant C(t-s)^{\frac{\alpha}{2}}\hat{p}_{\lambda'}(s,\boldsymbol{x};t,\boldsymbol{y}). \tag{3.22}$$

Proof. Note that by (2.5),

$$g_{\lambda/\kappa_3}\big(t-s, \boldsymbol{x} - \boldsymbol{\theta}_{s,t}(\boldsymbol{y})\big) \lesssim \hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}) \lesssim g_{\kappa_3\lambda}\big(t-s, \boldsymbol{x} - \boldsymbol{\theta}_{s,t}(\boldsymbol{y})\big).$$

The desired estimate now follows by definition. □

We also need the following convolution type inequality for the Gaussian functions $\hat{p}_\lambda$.

Lemma 3.4 (Reproduction property)**.** *Under (2.1) and (2.2), for any $T > 0$, there are constants $\kappa_7, C_3 \geqslant 1$ depending only on Θ_T such that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2d}$, $0 \leqslant s < r \leqslant t \leqslant T$ and $\lambda_1, \lambda_2 > 0$,*

$$C_3^{-1}\hat{p}_{(\lambda_1\wedge\lambda_2)/\kappa_7}(s,\boldsymbol{x};t,\boldsymbol{y}) \leqslant \int_{\mathbb{R}^{2d}} \hat{p}_{\lambda_1}(s,\boldsymbol{x};r,\boldsymbol{z})\hat{p}_{\lambda_2}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z} \leqslant C_3\hat{p}_{\kappa_7(\lambda_1\vee\lambda_2)}(s,\boldsymbol{x};t,\boldsymbol{y}). \tag{3.23}$$

Proof. Recalling (3.21) and (1.14), we first consider the convolution

$$I := \int_{\mathbb{R}^{2d}} g_{\lambda_1}(\epsilon_1, \boldsymbol{x}' - \boldsymbol{z}) g_{\lambda_2}(\epsilon_2, \boldsymbol{z} - \boldsymbol{x}'')\mathrm{d}\boldsymbol{z}.$$

Clearly, up to a constant dependent on d and λ_1, λ_2 , I is the density of the sum of two independent Gaussian vectors with means $\boldsymbol{x}'$ and $\boldsymbol{x}''$ respectively. Therefore I is a Gaussian function with mean $\boldsymbol{x}' - \boldsymbol{x}''$ and covariance matrix $\lambda_1 \mathbb{T}_{\epsilon_1}^{-2} + \lambda_2 \mathbb{T}_{\epsilon_2}^{-2}$. We have

$$\tfrac{\lambda_1 \wedge \lambda_2}{8} \mathbb{T}_{\epsilon_1+\epsilon_2}^{-2} \leqslant \lambda_1 \mathbb{T}_{\epsilon_1}^{-2} + \lambda_2 \mathbb{T}_{\epsilon_2}^{-2} \leqslant (\lambda_1 \vee \lambda_2) \mathbb{T}_{\epsilon_1+\epsilon_2}^{-2},$$

which yields

$$g_{(\lambda_1\wedge\lambda_2)/8}(\epsilon_1 + \epsilon_2, \boldsymbol{x}' - \boldsymbol{x}'') \lesssim I \lesssim g_{\lambda_1\vee\lambda_2}(\epsilon_1 + \epsilon_2, \boldsymbol{x}' - \boldsymbol{x}''). \tag{3.24}$$

To prove (3.23), it suffices to notice that by Lemma 4.1,

$$g_{\lambda_2/\kappa_3}(t - r, \boldsymbol{z} - \boldsymbol{\theta}_{r,t}(\boldsymbol{y})) \lesssim \hat{p}_{\lambda_2}(r, \boldsymbol{z}; t, \boldsymbol{y}) \lesssim g_{\kappa_3\lambda_2}(t - r, \boldsymbol{z} - \boldsymbol{\theta}_{r,t}(\boldsymbol{y})).$$

The claim then follows from (3.24) with $\boldsymbol{x}' = \boldsymbol{\theta}_{r,s}(\boldsymbol{x})$ and $\boldsymbol{x}'' = \boldsymbol{\theta}_{r,t}(\boldsymbol{y})$ and using (2.9). □

The following lemma is a direct consequence of Proposition 3.1 and Lemma 3.2.

Lemma 3.5. *Under* $(\mathbf{H}^0_\sigma)$, $(\mathbf{H}^0_{\mathbf{F}})$ *and* (3.1), *for any* $T > 0$ *and* $j = (j_1, j_2) \in \mathbb{N}_0^2$, *there are constants* $\lambda_j, C_j \geqslant 1$ *depending only on* Θ_T *such that for all* $0 \leqslant s < t \leqslant T$ *and* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2d}$,

$$C_0^{-1} \hat{p}_{\lambda_0^{-1}}\big(s, \boldsymbol{x}; t, \boldsymbol{y}\big) \leqslant \widetilde{p}_i(s, \boldsymbol{x}; t, \boldsymbol{y}) \leqslant C_0 \hat{p}_{\lambda_0}\big(s, \boldsymbol{x}; t, \boldsymbol{y}\big),\ \ i = 0, 1, \tag{3.25}$$

and

$$|\nabla^{j_1}_{x_1} \nabla^{j_2}_{x_2} \widetilde{p}^{(\tau,\boldsymbol{\xi})}(s, \boldsymbol{x}; t, \boldsymbol{y})|_{(\tau,\boldsymbol{\xi})=(s,\boldsymbol{x})\,\mathrm{or}\,(t,\boldsymbol{y})} \leqslant C_j (t - s)^{-\frac{j_1+3j_2}{2}} \hat{p}_{\lambda_j}\big(s, \boldsymbol{x}; t, \boldsymbol{y}\big). \tag{3.26}$$

Moreover, for any $\alpha \in [0, 1]$, *recalling* $\widetilde{p}_1(s, \boldsymbol{x}; t, \boldsymbol{y}) := \widetilde{p}^{(t,\boldsymbol{y})}(s, \boldsymbol{x}; t, \boldsymbol{y})$,

$$\begin{aligned} &|\nabla^{j_1}_{x_1} \nabla^{j_2}_{x_2} \big(\widetilde{p}_1(s, \boldsymbol{x}; t, \boldsymbol{y}) - \widetilde{p}_1(s, \boldsymbol{x}'; t, \boldsymbol{y})\big)| \\ &\qquad \leqslant C_j |\boldsymbol{x} - \boldsymbol{x}'|^\alpha_{\mathbf{d}} (t - s)^{-\frac{\alpha+j_1+3j_2}{2}} \big(\hat{p}_{\lambda_j}(s, \boldsymbol{x}; t, \boldsymbol{y}) + \hat{p}_{\lambda_j}(s, \boldsymbol{x}'; t, \boldsymbol{y})\big). \end{aligned} \tag{3.27}$$

Proof. Estimates (3.25) and (3.26) are direct by (3.14), (3.15), Lemma 3.2 and (2.16). Here we prove (3.27) with $j = (1, 0)$ for simplicity since the general case is analogous. The off-diagonal regime $|\boldsymbol{x} - \boldsymbol{x}'|_{\mathbf{d}} > (t - s)^{\frac{1}{2}}$ is straightforward from (3.26). Assume now $|\boldsymbol{x} - \boldsymbol{x}'|_{\mathbf{d}} \leqslant (t - s)^{\frac{1}{2}}$. Then we have

$$\begin{aligned} |\nabla_{x_1} \big(\widetilde{p}_1(s, \boldsymbol{x}; t, \boldsymbol{y}) - \widetilde{p}_1(s, \boldsymbol{x}'; t, \boldsymbol{y})\big)| &\leqslant \sum_{i=1}^2 |x_i - x_i'| \sup_{\eta\in[0,1]} |\nabla_{x_i} \nabla_{x_1} \widetilde{p}_1(s, \boldsymbol{x} + \eta(\boldsymbol{x}' - \boldsymbol{x}); t, \boldsymbol{y})| \\ &\lesssim \sum_{i=1}^2 |x_i - x_i'|(t - s)^{-i} \sup_{\eta\in[0,1]} \hat{p}_{\lambda_2}(s, \boldsymbol{x} + \eta(\boldsymbol{x}' - \boldsymbol{x}); t, \boldsymbol{y}) \\ &\lesssim |\boldsymbol{x} - \boldsymbol{x}'|^\alpha_{\mathbf{d}} (t - s)^{-\frac{\alpha+1}{2}} \hat{p}_{\lambda_j}(s, \boldsymbol{x}; t, \boldsymbol{y}), \end{aligned}$$

where in the last inequality we have used $|\boldsymbol{x} - \boldsymbol{x}'|_{\mathbf{d}} \leqslant (t - s)^{\frac{1}{2}}$ and (2.10). The proof is complete. □

3.2 Derivation of the upper bound

For notational convenience, we write from now on

$$\begin{aligned} H(s,\boldsymbol{x};t,\boldsymbol{y}) &:= (\mathcal{L}_{s,\boldsymbol{x}} - \widetilde{\mathcal{L}}^{(t,\boldsymbol{y})}_{s,\boldsymbol{x}})\widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y}) = (a(s,\boldsymbol{x}) - a(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{y}))) \cdot \nabla^2_{x_1}\widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y}) \\ &\quad + (F_1(s,\boldsymbol{x}) - F_1(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{y}))) \cdot \nabla_{x_1}\widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y}) + \mathcal{T}_{F_2(s)}(\boldsymbol{x},\boldsymbol{\theta}_{s,t}(\boldsymbol{y})) \cdot \nabla_{x_2}\widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y}), \end{aligned} \tag{3.28}$$

where $\mathcal{T}_{F_2(s)}$ is defined by (1.13), and

$$(p \otimes H)(s,\boldsymbol{x};t,\boldsymbol{y}) = \int_s^t \int_{\mathbb{R}^d} p(s,\boldsymbol{x};r,\boldsymbol{z})H(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r. \tag{3.29}$$

Thus, from the Duhamel representation (3.17), we have

$$p(s,\boldsymbol{x};t,\boldsymbol{y}) = \widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y}) + (p \otimes H)(s,\boldsymbol{x};t,\boldsymbol{y}). \tag{3.30}$$

For $N \geqslant 2$, iterating $N-1$-times the identity (3.30), we obtain

$$p(s,\boldsymbol{x};t,\boldsymbol{y}) = \widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y}) + \sum_{j=1}^{N-1} (\widetilde{p}_1 \otimes H^{\otimes j})(s,\boldsymbol{x};t,\boldsymbol{y}) + (p \otimes H^{\otimes N})(s,\boldsymbol{x};t,\boldsymbol{y}). \tag{3.31}$$

Proposition 3.6 (Control of the parametrix expansion)**.** *For any $T > 0$ and $N \in \mathbb{N}$, there are constants $C_N, \lambda_N > 0$ depending only on Θ_T such that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2d}$ and $0 \leqslant s < t \leqslant T$,*

$$|H^{\otimes N}(s,\boldsymbol{x};t,\boldsymbol{y})| \leqslant C_N (t-s)^{-1+\frac{N\gamma}{2}} \hat{p}_{\lambda_N}(s,\boldsymbol{x};t,\boldsymbol{y}), \tag{3.32}$$

where $\lambda_N \to \infty$ as $N \to \infty$. In particular,

$$p(s,\boldsymbol{x};t,\boldsymbol{y}) \leqslant C_{N-1}\hat{p}_{\lambda_{N-1}}(s,\boldsymbol{x};t,\boldsymbol{y}) + |(p \otimes H^{\otimes N})(s,\boldsymbol{x};t,\boldsymbol{y})|.$$

Proof. By (3.28), (3.15), (1.13) and (3.22), we have

$$\begin{aligned} |H(s,\boldsymbol{x};t,\boldsymbol{y})| &\leqslant |a(s,\boldsymbol{x}) - a(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{y}))| \cdot |\nabla^2_{x_1}\widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y})| \\ &\quad + |F_1(s,\boldsymbol{x}) - F_1(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{y}))| \cdot |\nabla_{x_1}\widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y})| \\ &\quad + |\mathcal{T}_{F_2(s)}(\boldsymbol{x},\boldsymbol{\theta}_{s,t}(\boldsymbol{y}))| \cdot |\nabla_{x_2}\widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y})| \\ &\lesssim (t-s)^{-1}|\boldsymbol{x} - \boldsymbol{\theta}_{s,t}(\boldsymbol{y})|^{\gamma}_{\mathbf{d}}\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}) \\ &\quad + (t-s)^{-\frac{1}{2}}(1 + |\boldsymbol{x} - \boldsymbol{\theta}_{s,t}(\boldsymbol{y})|)\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}) \\ &\quad + (t-s)^{-\frac{3}{2}}|\boldsymbol{x} - \boldsymbol{\theta}_{s,t}(\boldsymbol{y})|^{1+\gamma}_{\mathbf{d}}\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}) \\ &\lesssim (t-s)^{-1+\frac{\gamma}{2}}\hat{p}_{\lambda'}(s,\boldsymbol{x};t,\boldsymbol{y}). \end{aligned}$$

This gives the stated estimate for $N = 1$. For general $N \geqslant 2$, by induction, it is readily seen that:

$$\begin{aligned} |H^{\otimes N}(s,\boldsymbol{x};t,\boldsymbol{y})| &\lesssim \int_s^t (r-s)^{-1+\frac{(N-1)\gamma}{2}}(t-r)^{-1+\frac{\gamma}{2}} \left(\int_{\mathbb{R}^{2d}} \hat{p}_{\lambda_{N-1}}(s,\boldsymbol{x};r,\boldsymbol{z})\hat{p}_{\lambda_1}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}z \right) \mathrm{d}r \\ &\overset{(3.23)}{\lesssim} \left(\int_s^t (r-s)^{-1+\frac{(N-1)\gamma}{2}}(t-r)^{-1+\frac{\gamma}{2}}\mathrm{d}r \right) \hat{p}_{\lambda_N}(s,\boldsymbol{x};t,\boldsymbol{y}) \\ &\lesssim (t-s)^{-1+\frac{N\gamma}{2}}\hat{p}_{\lambda_N}(s,\boldsymbol{x};t,\boldsymbol{y}). \end{aligned}$$

This completes the proof. □

We carefully remark that we have to stop the iteration at some fixed N to avoid the explosion of λ_N as N goes to infinity. The following proposition provides a control for the remainder and concludes the proof of the upper bound.

Proposition 3.7 (Control of the remainder)**.** *Let N be large enough such that*

$$-1+\frac{N\gamma}{2}>2d.$$

Then there exist constants $C_0,\lambda_0>0$ such that for all $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^{2d}$ and $0\leqslant s<t\leqslant T$,

$$|(p\otimes H^{\otimes N})(s,\boldsymbol{x};t,\boldsymbol{y})|\leqslant C_0\hat{p}_{\lambda_0}(s,\boldsymbol{x};t,\boldsymbol{y}). \tag{3.33}$$

Proof. We first recall that, from the intrinsic scaling properties of SDE (1.3) we can restrict w.l.o.g to the case $s=0,t=1$ for the proof (we refer to [9] for additional details).

Indeed, recall first from (1.9) that for any $\lambda>0$, the *intrinsic scale matrix* writes:

$$\mathbb{T}_\lambda^{-1}:=\begin{pmatrix}\lambda^{-\frac{1}{2}}\mathbb{I}_{d\times d} & 0_{d\times d}\\ 0_{d\times d} & \lambda^{-\frac{3}{2}}\mathbb{I}_{d\times d}\end{pmatrix}.$$

Fix then $s\geqslant 0$ and $\lambda>0$. We define

$$\mathbf{X}_t^{\lambda,s}:=\mathbb{T}_\lambda^{-1}\mathbf{X}_{\lambda t+s},\ t\geqslant 0. \tag{3.34}$$

By the change of variables, it is easy to see that $(\mathbf{X}_t^{\lambda,s})_{t\geqslant 0}$ satisfies

$$\mathrm{d}\mathbf{X}_t^{\lambda,s}=\mathbf{F}^{\lambda,s}(t,\mathbf{X}_t^{\lambda,s})\mathrm{d}t+B\sigma^{\lambda,s}(t,\mathbf{X}_t^{\lambda,s})\mathrm{d}W_t^\lambda,$$

where $W_t^\lambda:=\lambda^{-\frac{1}{2}}W_{\lambda t}$ is a still Brownian motion, and

$$\mathbf{F}^{\lambda,s}(t,\boldsymbol{x}):=\lambda\mathbb{T}_\lambda^{-1}\mathbf{F}\big(s+\lambda t,\mathbb{T}_\lambda\boldsymbol{x}\big),\quad \sigma^{\lambda,s}(t,\boldsymbol{x}):=\sigma\big(s+\lambda t,\mathbb{T}_\lambda\boldsymbol{x}\big). \tag{3.35}$$

In particular, let $p(s,\boldsymbol{x};t,\boldsymbol{y})$ (resp. $p^{\lambda,s}(\boldsymbol{x};t,\boldsymbol{y})$) be the density of $\mathbf{X}_t$ (resp. $\mathbf{X}_t^{\lambda,s}$) starting from $\boldsymbol{x}$ at time s (resp. 0). Then

$$p(s,\boldsymbol{x};t,\boldsymbol{y})=\lambda^{2d}p^{\lambda,s}(\mathbb{T}_\lambda^{-1}\boldsymbol{x};\lambda^{-1}t,\mathbb{T}_\lambda^{-1}\boldsymbol{y}). \tag{3.36}$$

From the scaling property (3.36), it thus suffices to consider the case $(s,t)=(0,1)$. First of all, by (3.32), we have

$$\begin{aligned}\mathscr{I}:=|(p\otimes H^{\otimes N})(0,\boldsymbol{x};1,\boldsymbol{y})|&\lesssim\int_0^1(1-r)^{-1+\frac{N\gamma}{2}}\int_{\mathbb{R}^{2d}}p(0,\boldsymbol{x};r,\boldsymbol{z})\hat{p}_{\lambda_N}(r,\boldsymbol{z};1,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r\\&=\int_0^1(1-r)^{-1+\frac{N\gamma}{2}}\mathbb{E}\hat{p}_{\lambda_N}(r,\mathbf{X}_{r,0}(\boldsymbol{x});1,\boldsymbol{y})\mathrm{d}r.\end{aligned}$$

Noting that by Lemma 2.8 in [31],

$$\mathbb{E}\hat{p}_{\lambda_N}(r,\mathbf{X}_{r,0}(\boldsymbol{x});1,\boldsymbol{y})\leqslant C_1\sup_{\boldsymbol{z}\in\mathbb{R}^{2d}}\exp\Big\{\ln\hat{p}_{\lambda_N}(r,\boldsymbol{z};1,\boldsymbol{y})-C_2|\boldsymbol{z}-\boldsymbol{\theta}_{r,0}(\boldsymbol{x})|^2\Big\},$$

and

$$
\begin{aligned}
\ln \hat{p}_{\lambda_N}(r,\boldsymbol{z};1,\boldsymbol{y}) &\overset{(3.21)}{=} \ln(1-r)^{-2d} - \lambda_N|\mathbb{T}^{-1}_{1-r}(\boldsymbol{\theta}_{1,r}(\boldsymbol{z})-\boldsymbol{y})|^2 \\
&\overset{(2.5)}{\leqslant} \ln(1-r)^{-2d} - \lambda'_N|\mathbb{T}^{-1}_{1-r}(\boldsymbol{z}-\boldsymbol{\theta}_{r,1}(\boldsymbol{y}))|^2 + C_3 \\
&\leqslant \ln(1-r)^{-2d} - \lambda'_N|\boldsymbol{z}-\boldsymbol{\theta}_{r,1}(\boldsymbol{y})|^2 + C_3,
\end{aligned}
$$

we further have

$$
\begin{aligned}
\mathbb{E}\hat{p}_{\lambda_N}(r,\mathbf{X}_{r,0}(\boldsymbol{x});1,\boldsymbol{y}) &\lesssim (1-r)^{-2d}\exp\left\{-\lambda''_N \inf_{\boldsymbol{z}\in\mathbb{R}^{2d}}\left\{|\boldsymbol{z}-\boldsymbol{\theta}_{r,1}(\boldsymbol{y})|^2 + |\boldsymbol{z}-\boldsymbol{\theta}_{r,0}(\boldsymbol{x})|^2\right\}\right\} \\
&= (1-r)^{-2d}\exp\left\{-\frac{\lambda''_N}{2}|\boldsymbol{\theta}_{r,1}(\boldsymbol{y})-\boldsymbol{\theta}_{r,0}(\boldsymbol{x})|^2\right\} \\
&\overset{(2.5)}{\leqslant} (1-r)^{-2d}\exp\left\{-\lambda'''_N|\boldsymbol{\theta}_{1,0}(\boldsymbol{x})-\boldsymbol{y}|^2\right\}.
\end{aligned}
$$

Hence,

$$
\mathscr{I} \lesssim \left(\int_0^1 (1-r)^{-1+\frac{N\gamma}{2}-2d}\mathrm{d}r\right)\mathrm{e}^{-\lambda'''_N|\boldsymbol{\theta}_{1,0}(\boldsymbol{x})-\boldsymbol{y}|^2} \lesssim \mathrm{e}^{-\lambda'''_N|\boldsymbol{\theta}_{1,0}(\boldsymbol{x})-\boldsymbol{y}|^2}.
$$

The proof is complete. □

3.3 Derivation of the lower bound

We first derive a local bound, starting from the one step parametrix expansion (3.30): by (3.15) and the upper bound for p, we have

$$
\begin{aligned}
p(s,\boldsymbol{x};t,\boldsymbol{y}) &= \widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y}) + (p\otimes H)(s,\boldsymbol{x};t,\boldsymbol{y}) \\
&\geqslant c_0\hat{p}_{\lambda_0^{-1}}(s,\boldsymbol{x};t,\boldsymbol{y}) - \int_s^t\int_{\mathbb{R}^{2d}} p(s,\boldsymbol{x};r,\boldsymbol{z})|H(r,\boldsymbol{z};t,\boldsymbol{y})|\mathrm{d}\boldsymbol{z}\mathrm{d}r \\
&\geqslant c_0\hat{p}_{\lambda_0^{-1}}(s,\boldsymbol{x};t,\boldsymbol{y}) - C\int_s^t\int_{\mathbb{R}^{2d}} \hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})(t-r)^{\frac{\gamma}{2}-1}\hat{p}_\lambda(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r \\
&\overset{(3.23)}{\geqslant} C^{-1}\hat{p}_{\lambda^{-1}}(s,\boldsymbol{x};t,\boldsymbol{y}) - C(t-s)^{\frac{\gamma}{2}}\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}),
\end{aligned}
$$

for sufficiently large λ and C, depending on Θ_T. Suppose that

$$
|\mathbb{T}^{-1}_{t-s}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x})-\boldsymbol{y})| \leqslant \kappa_5(2\kappa_6+1)+2\kappa_3+1 =: \kappa_9,
$$

where κ_3,κ_5 and κ_6 are from (2.5), (2.20) and (2.21). Then

$$
p(s,\boldsymbol{x};t,\boldsymbol{y}) \geqslant (t-s)^{-2d}(C^{-1}\mathrm{e}^{-\lambda\kappa_9^2} - C(t-s)^{\frac{\gamma}{2}}) \geqslant \frac{\mathrm{e}^{-\lambda\kappa_9^2}}{2C}(t-s)^{-2d} = C_0(t-s)^{-2d}, \tag{3.37}
$$

provided that

$$
t-s \leqslant (2C^2)^{-\frac{2}{\gamma}}\mathrm{e}^{-\frac{2\lambda\kappa_9^2}{\gamma}} =: T_0.
$$

Now for fixed $0 \leqslant s < t \leqslant T$ with $t - s \leqslant T_0$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2d}$, we assume

$$|\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}) - \boldsymbol{y})| \geqslant \kappa_9.$$

Let M be the smallest integer such that

$$M - 1 \leqslant |\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}) - \boldsymbol{y})|^2 < M. \tag{3.38}$$

Define

$$\delta := \tfrac{t-s}{M}, \quad t_j := s + j\delta, \quad j = 0, 1, \cdots, M.$$

Let $\boldsymbol{\phi}_{t,s}(\boldsymbol{x})$ be the optimal curve in Proposition 2.5 so that the corresponding control φ has the estimate

$$\sup_{r\in[s,t]} |\varphi_r| \leqslant \kappa_6\big(|\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}) - \boldsymbol{y})| + 1\big)/\sqrt{t-s} \leqslant \kappa_6(\sqrt{M} + 1)/\sqrt{t-s}.$$

Define

$$\boldsymbol{\xi}_j := \boldsymbol{\phi}_{t_j,s}(\boldsymbol{x}), \quad j = 0, 1, \cdots, M.$$

For any $j = 0, \cdots, M-1$, by Proposition 2.5, we have

$$\begin{aligned} |\mathbb{T}_\delta^{-1}\big(\boldsymbol{\theta}_{t_{j+1},t_j}(\boldsymbol{\xi}_j) - \boldsymbol{\xi}_{j+1}\big)| &\lesssim_{\kappa_5} I(t_j, \boldsymbol{\xi}_j; t_{j+1}, \boldsymbol{\xi}_{j+1}) + 1 \leqslant \left(\int_{t_j}^{t_{j+1}} |\varphi_r|^2 \mathrm{d}r\right)^{1/2} + 1 \\ &\leqslant (t_{j+1} - t_j)^{1/2} \sup_{r\in[s,t]} |\varphi_r| + 1 \leqslant 2\kappa_6 + 1, \end{aligned} \tag{3.39}$$

from the very definition of δ and the previous control on $\sup_{r\in[s,t]} |\varphi_r|$. Now set

$$\Sigma_0 := \{\boldsymbol{\xi}_0\} = \{\boldsymbol{x}\}, \quad \Sigma_M := \{\boldsymbol{\xi}_M\} = \{\boldsymbol{y}\},$$

and for $j = 1, \cdots, M-1$,

$$\Sigma_j := \Big\{\boldsymbol{z} \in \mathbb{R}^{2d} : |\mathbb{T}_\delta^{-1}(\boldsymbol{z} - \boldsymbol{\xi}_j)| \leqslant 1\Big\}.$$

By (3.39) and (2.5), for any $j = 0, \cdots, M-1$, we have that for $\boldsymbol{z}_j \in \Sigma_j$ and $\boldsymbol{z}_{j+1} \in \Sigma_{j+1}$,

$$\begin{aligned} |\mathbb{T}_\delta^{-1}(\boldsymbol{\theta}_{t_{j+1},t_j}(\boldsymbol{z}_j) - \boldsymbol{z}_{j+1})| &\leqslant |\mathbb{T}_\delta^{-1}(\boldsymbol{\theta}_{t_{j+1},t_j}(\boldsymbol{\xi}_j) - \boldsymbol{\xi}_{j+1})| + |\mathbb{T}_\delta^{-1}(\boldsymbol{z}_{j+1} - \boldsymbol{\xi}_{j+1})| \\ &\quad + |\mathbb{T}_\delta^{-1}(\boldsymbol{\theta}_{t_{j+1},t_j}(\boldsymbol{z}_j) - \boldsymbol{\theta}_{t_{j+1},t_j}(\boldsymbol{\xi}_j))| \\ &\leqslant |\mathbb{T}_\delta^{-1}(\boldsymbol{\theta}_{t_{j+1},t_j}(\boldsymbol{\xi}_j) - \boldsymbol{\xi}_{j+1})| + |\mathbb{T}_\delta^{-1}(\boldsymbol{z}_{j+1} - \boldsymbol{\xi}_{j+1})| \\ &\quad + \kappa_3(|\mathbb{T}_\delta^{-1}(\boldsymbol{z}_j - \boldsymbol{\xi}_j)| + 1) \\ &\leqslant \kappa_5(2\kappa_6 + 1) + 2\kappa_3 + 1 = \kappa_9. \end{aligned}$$

This precisely means that the previous diagonal lower bound holds for $p(t_j, \boldsymbol{z}_j; t_{j+1}, \boldsymbol{z}_{j+1})$. Thus, by the Chapman-Kolmogorov equation and (3.37), we have

$$\begin{aligned} p(s, \boldsymbol{x}; t, \boldsymbol{y}) &= \int_{\mathbb{R}^{2d}} \cdots \int_{\mathbb{R}^{2d}} p(t_0, \boldsymbol{x}; u, \boldsymbol{z}_1) \cdots p(t_{M-1}, \boldsymbol{z}_{M-1}; t_M, \boldsymbol{y}) \mathrm{d}\boldsymbol{z}_1 \cdots \mathrm{d}\boldsymbol{z}_{M-1} \\ &\geqslant \int_{\Sigma_1} \cdots \int_{\Sigma_{M-1}} p(t_0, \boldsymbol{z}_0; u, \boldsymbol{z}_1) \cdots p(t_{M-1}, \boldsymbol{z}_{M-1}; t_M, \boldsymbol{z}_M) \mathrm{d}\boldsymbol{z}_1 \cdots \mathrm{d}\boldsymbol{z}_{M-1} \end{aligned}$$

$$\geqslant (C_0\delta^{-2d})^M \int_{\Sigma_1} \cdots \int_{\Sigma_{M-1}} \mathrm{d}\boldsymbol{z}_1 \cdots \mathrm{d}\boldsymbol{z}_{M-1} = (C_0\delta^{-2d})^M (C_4\delta^{2d})^{M-1},$$

where C_0 is given in (3.37), and the last equality is due to $|\Sigma_j| = C_4\delta^{2d}$ for some C_4 only depending on d. Recalling $\delta = (t-s)/M$ and M given in (3.38), we finally have

$$\begin{aligned} p(s,\boldsymbol{x};t,\boldsymbol{y}) &\geqslant C_0^M (C_4)^{M-1}\delta^{-2d} = (t-s)^{-2d} M^{2d} \exp\{M\log(C_0C_4)\}/C_4 \\ &\geqslant C_5(t-s)^{-2d}\exp\{-\lambda_0|\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x})-\boldsymbol{y})|^2\}. \end{aligned}$$

The lower bound is thus obtained for $t-s \leqslant T_0$ and all $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^{2d}$. For general $0 \leqslant s < t \leqslant T$, it again follows from the Chapman-Kolmogorov equation.

3.4 The parametrix series expansion of $p(s,\boldsymbol{x};t,\boldsymbol{y})$

We introduce, for $\delta > 0$ the SDE (1.1) with diffusion coefficient $\sigma(s,\boldsymbol{x}) = \delta\mathbb{I}_{d\times d}$ and denote with $\bar{p}_\delta$ the corresponding density. By scaling and the two-sided density bounds of Sections 3.2 and 3.3 it holds that, for any $\lambda > 0$, there exists $\delta = \delta(\lambda)$ large enough, and $C_\delta \geqslant 1$, $\lambda' > \lambda$ such that, for all $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^{2d}$ and $0 \leqslant s < t \leqslant T$,

$$C_\delta^{-1}\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}) \leqslant \bar{p}_\delta(s,\boldsymbol{x};t,\boldsymbol{y}) \leqslant C_\delta \hat{p}_{\lambda'}(s,\boldsymbol{x};t,\boldsymbol{y}). \tag{3.40}$$

By Lemmas 3.3 and 3.5, we may choose λ, and then $\delta = \delta(\lambda)$ such that, for any $j = (j_1,j_2) \in \mathbb{N}_0^2$ with $j_1 + 3j_2 \leqslant 3$, $k \in \{1,2\}$, $\beta \in [0,2]$ and for all $0 \leqslant s < t \leqslant T$, $\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y} \in \mathbb{R}^{2d}$ with $|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}} \leqslant (t-s)^{\frac{1}{2}}$,

$$|(\boldsymbol{x}-\boldsymbol{\theta}_{s,t}(\boldsymbol{y}))_k|^\beta |\nabla_{x_1}^{j_1}\nabla_{x_2}^{j_2}\widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y})| \lesssim (t-s)^{\beta(k-\frac{1}{2})-\frac{k_1+3j_2}{2}}\bar{p}_\delta(s,\boldsymbol{x};t,\boldsymbol{y}), \tag{3.41}$$

and also, for any $\alpha \in [0,1]$,

$$\begin{aligned} &|(\boldsymbol{x}-\boldsymbol{\theta}_{s,t}(\boldsymbol{y}))_k|^\beta |\nabla_{x_1}^{j_1}\nabla_{x_2}^{j_2}(\widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y}) - \widetilde{p}_1(s,\boldsymbol{x}';t,\boldsymbol{y}))| \\ &\qquad \lesssim |\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^\alpha (t-s)^{\beta(k-\frac{1}{2})-\frac{\alpha+j_1+3j_2}{2}}\bar{p}_\delta(s,\boldsymbol{x};t,\boldsymbol{y}). \end{aligned} \tag{3.42}$$

We shall fix from now on δ such that (3.41) and (3.42) hold and for simplicity we write $\bar{p} = \bar{p}_\delta$. Importantly, $\bar{p}$ enjoys the Chapman-Kolmogorov property, namely, for all $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^{2d}$ and $0 \leqslant t < r < s \leqslant T$,

$$\int_{\mathbb{R}^{2d}} \bar{p}(s,\boldsymbol{x};r,\boldsymbol{z})\bar{p}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z} = \bar{p}(s,\boldsymbol{x};t,\boldsymbol{y}). \tag{3.43}$$

Proposition 3.8. *Under* $(\mathbf{H}_\sigma^\gamma)$, $(\mathbf{H}_\mathbf{F}^\gamma)$ *and* (3.1), *the density* $p(s,\boldsymbol{x};t,\boldsymbol{y})$ *admits the following parametrix expansion*

$$p(s,\boldsymbol{x};t,\boldsymbol{y}) = \widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y}) + \widetilde{p}_1 \otimes \mathcal{H}(s,\boldsymbol{x};t,\boldsymbol{y}), \tag{3.44}$$

where $\mathcal{H} := \sum_{k\geqslant 1} H^{\otimes k}$ *enjoys the following estimates: for all* $0 \leqslant s < t \leqslant T$ *and* $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^{2d}$,

$$|\mathcal{H}(s,\boldsymbol{x};t,\boldsymbol{y})| \lesssim (t-s)^{\frac{\gamma}{2}-1}\bar{p}(s,\boldsymbol{x};t,\boldsymbol{y}). \tag{3.45}$$

If in addition F_1 *also satisfies* (1.23), *then for all* $\varepsilon \in (0,\gamma)$ *and* $\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y} \in \mathbb{R}^{2d}$,

$$|\mathcal{H}(s,\boldsymbol{x};t,\boldsymbol{y}) - \mathcal{H}(s,\boldsymbol{x}';t,\boldsymbol{y})| \lesssim |\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\gamma-\varepsilon}(t-s)^{\frac{\varepsilon}{2}-1}\left(\bar{p}(s,\boldsymbol{x};t,\boldsymbol{y}) + \bar{p}(s,\boldsymbol{x}';t,\boldsymbol{y})\right). \tag{3.46}$$

Proof. Let us first prove (3.45) and (3.46). From the definition (3.28) of H, the proof of Proposition 3.6 and (3.41), we derive that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2d}$ and $0 \leqslant s < t \leqslant T$,

$$|H(s, \boldsymbol{x}; t, \boldsymbol{y})| \leqslant C(t-s)^{\frac{\gamma}{2}-1} \bar{p}(s, \boldsymbol{x}; t, \boldsymbol{y}). \tag{3.47}$$

The point here is that $\bar{p}$ is a true density and enjoys the reproduction property (3.43). This allows to manage the iteration procedure without deteriorating the constants at every step. Indeed by direct induction, for $k \geqslant 1$,

$$|H^{\otimes k}(s, \boldsymbol{x}; t, \boldsymbol{y})| \leqslant C^k \frac{\Gamma^k(\gamma/2)}{\Gamma(k\gamma/2)} (t-s)^{\frac{k\gamma}{2}-1} \bar{p}(s, \boldsymbol{x}; t, \boldsymbol{y}), \tag{3.48}$$

where Γ is the Euler-Gamma function. Estimate (3.45) easily follows.

Next let us consider (3.46). When $|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}} > (t-s)^{\frac{1}{2}}$ (off-diagonal regime) the estimate directly follows from (3.45). When $|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}} \leqslant (t-s)^{\frac{1}{2}}$ we restart from the first term of the expansion. By (3.28), we have

$$\begin{aligned}
&|H(s, \boldsymbol{x}; t, \boldsymbol{y}) - H(s, \boldsymbol{x}'; t, \boldsymbol{y})| \\
&\quad = \Big|(\mathcal{L}_{s,\boldsymbol{x}} - \widetilde{\mathcal{L}}^{(t,\boldsymbol{y})}_{s,\boldsymbol{x}})\widetilde{p}_1(s, \boldsymbol{x}; t, \boldsymbol{y}) - (\mathcal{L}_{s,\boldsymbol{x}} - \widetilde{\mathcal{L}}^{(t,\boldsymbol{y})}_{s,\boldsymbol{x}})\widetilde{p}_1(s, \boldsymbol{x}'; t, \boldsymbol{y})\Big| \\
&\quad \leqslant \Big[|a(s,\boldsymbol{x}) - a(s, \boldsymbol{\theta}_{s,t}(\boldsymbol{y}))||\nabla^2_{x_1}\left(\widetilde{p}_1(s, \boldsymbol{x}; t, \boldsymbol{y}) - \widetilde{p}_1(s, \boldsymbol{x}'; t, \boldsymbol{y})\right)| \\
&\qquad + |a(s,\boldsymbol{x}) - a(s,\boldsymbol{x}')||\nabla^2_{x_1}\widetilde{p}_1(s, \boldsymbol{x}'; t, \boldsymbol{y})|\Big] \\
&\qquad + \Big[|F_1(s,\boldsymbol{x}) - F_1(s, \boldsymbol{\theta}_{s,t}(\boldsymbol{y}))||\nabla_{x_1}\left(\widetilde{p}_1(s, \boldsymbol{x}; t, \boldsymbol{y}) - \widetilde{p}_1(s, \boldsymbol{x}'; t, \boldsymbol{y})\right)| \\
&\qquad + |F_1(s,\boldsymbol{x}) - F_1(s,\boldsymbol{x}')||\nabla_{x_1}\widetilde{p}_1(s, \boldsymbol{x}'; t, \boldsymbol{y})|\Big] \\
&\qquad + \Big[|\mathcal{T}_{F_2(s)}(\boldsymbol{x}, \boldsymbol{\theta}_{s,t}(\boldsymbol{y}))|\,|\nabla_{x_2}\left(\widetilde{p}_1(s, \boldsymbol{x}; t, \boldsymbol{y}) - \widetilde{p}_1(s, \boldsymbol{x}'; t, \boldsymbol{y})\right)| \\
&\qquad + |\mathcal{T}_{F_2(s)}(\boldsymbol{x}, \boldsymbol{\theta}_{s,t}(\boldsymbol{y})) - \mathcal{T}_{F_2(s)}(\boldsymbol{x}', \boldsymbol{\theta}_{s,t}(\boldsymbol{y}))|\,|\nabla_{x_2}\widetilde{p}_1(s, \boldsymbol{x}'; t, \boldsymbol{y})|\Big] \\
&\quad =: I_1 + I_2 + I_3.
\end{aligned}$$

By (3.41), (3.42) and using that $|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}} \leqslant (t-s)^{\frac{1}{2}}$ we have

$$\begin{aligned}
I_1 &\lesssim |\boldsymbol{x}-\boldsymbol{x}'|^{\gamma}_{\mathbf{d}} \left(\frac{|\boldsymbol{x} - \boldsymbol{\theta}_{s,t}(\boldsymbol{y})|^{\gamma}_{\mathbf{d}}}{(t-s)^{1+\frac{\gamma}{2}}} \hat{p}_{\lambda_2}(s, \boldsymbol{x}; t, \boldsymbol{y}) + \frac{1}{t-s} \hat{p}_{\lambda_2}(s, \boldsymbol{x}'; t, \boldsymbol{y}) \right) \\
&\lesssim |\boldsymbol{x}-\boldsymbol{x}'|^{\gamma-\varepsilon}_{\mathbf{d}} (t-s)^{\frac{\varepsilon}{2}-1} \left(\bar{p}(s, \boldsymbol{x}; t, \boldsymbol{y}) + \bar{p}(s, \boldsymbol{x}'; t, \boldsymbol{y})\right).
\end{aligned}$$

Similarly, by (1.19) we also have

$$I_2 \lesssim |\boldsymbol{x}-\boldsymbol{x}'|^{\gamma}_{\mathbf{d}} (t-s)^{-\frac{1}{2}} \left(\bar{p}(s, \boldsymbol{x}; t, \boldsymbol{y}) + \bar{p}(s, \boldsymbol{x}'; t, \boldsymbol{y})\right),$$

and by (1.13), (3.41) and (3.42),

$$I_3 \lesssim |\boldsymbol{x} - \boldsymbol{\theta}_{s,t}(\boldsymbol{y})|^{1+\gamma}_{\mathbf{d}} (t-s)^{-\frac{3+\gamma}{2}} |\boldsymbol{x}-\boldsymbol{x}'|^{\gamma}_{\mathbf{d}} \hat{p}_{\lambda_3}(s, \boldsymbol{x}; t, \boldsymbol{y})$$

$$
\begin{aligned}
&+|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{1+\gamma}(t-s)^{-\frac{3}{2}}\hat{p}_{\lambda_3}(s,\boldsymbol{x}';t,\boldsymbol{y})\\
&\lesssim |\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\gamma-\varepsilon}(t-s)^{\frac{\varepsilon}{2}-1}\left(\bar{p}(s,\boldsymbol{x};t,\boldsymbol{y})+\bar{p}(s,\boldsymbol{x}';t,\boldsymbol{y})\right).
\end{aligned}
$$

Finally, we have

$$
\begin{aligned}
&|\mathcal{H}(s,\boldsymbol{x};t,\boldsymbol{y})-\mathcal{H}(s,\boldsymbol{x}';t,\boldsymbol{y})|\\
&\quad\leqslant |H(s,\boldsymbol{x};t,\boldsymbol{y})-H(s,\boldsymbol{x}';t,\boldsymbol{y})|+\int_s^t\int_{\mathbb{R}^{2d}}|H(s,\boldsymbol{x};r,\boldsymbol{z})-H(s,\boldsymbol{x}';r,\boldsymbol{z})||\mathcal{H}(r,\boldsymbol{z};t,\boldsymbol{y})|\mathrm{d}\boldsymbol{z}\mathrm{d}r\\
&\quad\lesssim |\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\gamma-\varepsilon}(t-s)^{\frac{\varepsilon}{2}-1}\left(\bar{p}(s,\boldsymbol{x};t,\boldsymbol{y})+\bar{p}(s,\boldsymbol{x}';t,\boldsymbol{y})\right)\\
&\qquad+\int_s^t\frac{|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\gamma-\frac{\varepsilon}{2}}}{(r-s)^{1-\frac{\varepsilon}{2}}(t-r)^{1-\frac{\varepsilon}{2}}}\int_{\mathbb{R}^{2d}}\left(\bar{p}(s,\boldsymbol{x};r,\boldsymbol{z})+\bar{p}(s,\boldsymbol{x}';r,\boldsymbol{z})\right)\bar{p}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r\\
&\quad\lesssim |\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\gamma-\varepsilon}(t-s)^{\frac{\varepsilon}{2}-1}\left(\bar{p}(s,\boldsymbol{x};t,\boldsymbol{y})+\bar{p}(s,\boldsymbol{x}';t,\boldsymbol{y})\right).
\end{aligned}
$$

The parametrix expansion (3.44) follows by letting $N\longrightarrow\infty$ in (3.31) and proving that the remainder $R^N:=p\otimes H^{\otimes N}$ converges to zero uniformly in every variable. Let $M_N:=C^N\Gamma^N\left(\frac{\gamma}{2}\right)\Gamma\left(\frac{N\gamma}{2}\right)^{-1}$ with C as in (3.48). By (3.48) we have

$$
\begin{aligned}
|(p\otimes H^{\otimes N})(s,\boldsymbol{x};t,\boldsymbol{y})|&\leqslant CM_N\int_s^t(t-r)^{\frac{N\gamma}{2}-1}\int_{\mathbb{R}^{2d}}\bar{p}(s,\boldsymbol{x};r,\boldsymbol{z})\bar{p}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r\\
&\leqslant C\frac{M_N}{N}(t-s)^{\frac{N\gamma}{2}}\bar{p}(s,\boldsymbol{x};t,\boldsymbol{y})\longrightarrow 0.
\end{aligned}
$$

The proof is completed. □

3.5 Sensitivities of the frozen densities with respect to spatial parameters

We provide here a lemma which quantifies the sensitivity of the frozen densities w.r.t. different (spatial) freezing parameters.

Lemma 3.9 (Sensitivity of the semigroup w.r.t. the freezing parameters)**.** *Under* $(\mathbf{H}^\gamma_\sigma)$, $(\mathbf{H}^\gamma_{\mathbf{F}})$ *and* (3.1)*, for any* $T>0$ *and* $j=(j_1,j_2)\in\mathbb{N}_0^2$*, there exist constants* $\lambda_j,C_j\geqslant 1$ *depending only on* Θ_T *such that for all* $0\leqslant s<t\leqslant T$ *and* $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^{2d}$*,*

$$
\left|\nabla_{x_1}^{j_1}\nabla_{x_2}^{j_2}\left(\widetilde{p}_1-\widetilde{p}^{(\tau,\boldsymbol{\xi})}\right)(s,\boldsymbol{x};t,\boldsymbol{y})\right|_{(\tau,\boldsymbol{\xi})=(s,\boldsymbol{x})}\leqslant C_j(t-s)^{\frac{\gamma-(j_1+3j_2)}{2}}\hat{p}_{\lambda_j}(s,\boldsymbol{x};t,\boldsymbol{y}); \tag{3.49}
$$

Moreover, for any $\boldsymbol{x}'\in\mathbb{R}^{2d}$

$$
\left|\nabla_{x_1}^{j_1}\nabla_{x_2}^{j_2}\left(\widetilde{p}^{(\tau,\boldsymbol{\xi})}-\widetilde{p}^{(\tau,\boldsymbol{\xi}')}\right)(s,\boldsymbol{x};t,\boldsymbol{y})\right|_{(\tau,\boldsymbol{\xi},\boldsymbol{\xi}')=(s,\boldsymbol{x},\boldsymbol{x}')}\leqslant C_j|\boldsymbol{x}-\boldsymbol{x}'|^{\gamma}(t-s)^{\frac{-(j_1+3j_2)}{2}}\hat{p}_{\lambda_j}(s,\boldsymbol{x};t,\boldsymbol{y}). \tag{3.50}
$$

Proof. We only prove (3.49) for $j_1=j_2=0$. For general $j_1,j_2\in\mathbb{N}_0$, the statement follows by the chain rule and tedious but similar calculations. The control (3.50) is derived analogously. For notational simplicity, we introduce

$$
\mathcal{C}_1=\mathbf{K}_{t,s}^{(s,\boldsymbol{x})},\ \mathcal{C}_2=\mathbf{K}_{t,s}^{(t,\boldsymbol{y})},\ \mathrm{w}_1=\boldsymbol{\theta}_{t,s}(\boldsymbol{x})-\boldsymbol{y},\ \mathrm{w}_2=\boldsymbol{\vartheta}_{t,s}^{(t,\boldsymbol{y})}(\boldsymbol{x})-\boldsymbol{y},
$$

and

$$\mathcal{A} := |\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\boldsymbol{x}) - \boldsymbol{y})| + 1.$$

Using the above notations and by definition (3.10), we have

$$\begin{aligned}
&\big|\big(\widetilde{p}_1 - \widetilde{p}^{(\tau,\boldsymbol{\xi})}\big)(s,\boldsymbol{x};t,\boldsymbol{y})\big|_{(\tau,\boldsymbol{\xi})=(s,\boldsymbol{x})} \\
&\quad = (2\pi)^{-d}\Big|\det(\mathcal{C}_2)^{-\frac{1}{2}}\exp\Big(-\tfrac{1}{2}\big|\mathcal{C}_2^{-\frac{1}{2}}\mathbf{w}_2\big|^2\Big) - \det(\mathcal{C}_1)^{-\frac{1}{2}}\exp\Big(-\tfrac{1}{2}\big|\mathcal{C}_1^{-\frac{1}{2}}\mathbf{w}_1\big|^2\Big)\Big|.
\end{aligned}$$

To show (3.49), it suffices to prove that for some $\lambda > 0$,

$$|(\det\mathcal{C}_1)^{-\frac{1}{2}} - (\det\mathcal{C}_2)^{-\frac{1}{2}}| \lesssim (t-s)^{-2d+\frac{\gamma}{2}}\mathcal{A}^\gamma, \tag{3.51}$$

$$\Big|\exp\big(-\tfrac{1}{2}|\mathcal{C}_1^{-\frac{1}{2}}\mathbf{w}_1|^2\big) - \exp\big(-\tfrac{1}{2}|\mathcal{C}_2^{-\frac{1}{2}}\mathbf{w}_2|\big)\Big| \lesssim (t-s)^{\frac{\gamma}{2}}\exp\big(-\lambda\mathcal{A}^2\big). \tag{3.52}$$

Note that by (2.17),

$$\mathcal{C}_i = \mathbb{T}_{t-s}\hat{\mathcal{C}}_i\mathbb{T}_{t-s}, \tag{3.53}$$

where

$$\hat{\mathcal{C}}_1 := \int_0^1 \hat{\mathbf{R}}_{1,r}^{(s,\boldsymbol{x})}\hat{\boldsymbol{\Sigma}}_r^{(s,\boldsymbol{x})}(\hat{\mathbf{R}}_{1,r}^{(s,\boldsymbol{x})}\hat{\boldsymbol{\Sigma}}_r^{(s,\boldsymbol{x})})^*\mathrm{d}r,\ \hat{\mathcal{C}}_2 := \int_0^1 \hat{\mathbf{R}}_{1,r}^{(t,\boldsymbol{y})}\hat{\boldsymbol{\Sigma}}_r^{(t,\boldsymbol{y})}(\hat{\mathbf{R}}_{1,r}^{(t,\boldsymbol{y})}\hat{\boldsymbol{\Sigma}}_r^{(t,\boldsymbol{y})})^*\mathrm{d}r,$$

and for $\eta(r) := s + (t-s)r$,

$$\hat{\boldsymbol{\Sigma}}_r^{(\tau,\boldsymbol{\xi})} := B\sigma(\eta(r), \boldsymbol{\theta}_{\eta(r),\tau}(\boldsymbol{\xi})),$$

and

$$\hat{\mathbf{R}}_{1,r}^{(\tau,\boldsymbol{\xi})} := \begin{pmatrix} \mathbb{I}_{d\times d} & 0_{d\times d} \\ \int_r^1 \nabla_{x_1}F_2(\eta(u), \boldsymbol{\theta}_{\eta(u),\tau}(\boldsymbol{\xi}))\mathrm{d}u & \mathbb{I}_{d\times d}\end{pmatrix}.$$

Since $\sigma, \nabla_{x_1}F_2 \in \mathscr{C}_{\mathbf{d}}^\gamma$, by (2.9) we have for any $r \in [0,1]$,

$$|\hat{\boldsymbol{\Sigma}}_r^{(s,\boldsymbol{x})} - \hat{\boldsymbol{\Sigma}}_r^{(t,\boldsymbol{y})}| \lesssim |\boldsymbol{\theta}_{\eta(r),s}(\boldsymbol{x}) - \boldsymbol{\theta}_{\eta(r),t}(\boldsymbol{y})|_{\mathbf{d}}^\gamma \lesssim (t-s)^{\frac{\gamma}{2}}\mathcal{A}^\gamma,$$

and

$$|\hat{\mathbf{R}}_{1,r}^{(s,\boldsymbol{x})} - \hat{\mathbf{R}}_{1,r}^{(t,\boldsymbol{y})}| \lesssim (t-s)^{\frac{\gamma}{2}}\mathcal{A}^\gamma.$$

Hence, by (2.15),

$$|\hat{\mathcal{C}}_1 - \hat{\mathcal{C}}_2| \lesssim (t-s)^{\frac{\gamma}{2}}\mathcal{A}^\gamma, \quad |\hat{\mathcal{C}}_i| \lesssim 1, \quad \det\hat{\mathcal{C}}_i \asymp 1, \tag{3.54}$$

and

$$\begin{aligned}
|(\det\mathcal{C}_1)^{-\frac{1}{2}} - (\det\mathcal{C}_1)^{-\frac{1}{2}}| &\lesssim (t-s)^{-2d}|(\det\hat{\mathcal{C}}_1)^{-\frac{1}{2}} - (\det\hat{\mathcal{C}}_1)^{-\frac{1}{2}}| \\
&\lesssim (t-s)^{-2d}|\det\hat{\mathcal{C}}_1 - \det\hat{\mathcal{C}}_1| \\
&\lesssim (t-s)^{-2d}|\hat{\mathcal{C}}_1 - \hat{\mathcal{C}}_2| \lesssim (t-s)^{-2d+\frac{\gamma}{2}}\mathcal{A}^\gamma.
\end{aligned}$$

Thus we obtain (3.51). For proving (3.52), without loss of generality, we may assume

$$|\mathcal{C}_1^{-\frac{1}{2}}\mathbf{w}_1| \leqslant |\mathcal{C}_2^{-\frac{1}{2}}\mathbf{w}_2|.$$

Then by $1-\mathrm{e}^{-x}\leqslant x$, we have

$$\Big|\exp\big(-\tfrac{1}{2}|\mathcal{C}_1^{-\frac{1}{2}}\mathbf{w}_1|^2\big)-\exp\big(-\tfrac{1}{2}|\mathcal{C}_2^{-\frac{1}{2}}\mathbf{w}_2|^2\big)\Big|\leqslant\tfrac{1}{2}(|\mathcal{C}_2^{-\frac{1}{2}}\mathbf{w}_2|^2-|\mathcal{C}_1^{-\frac{1}{2}}\mathbf{w}_1|^2)\exp\big(-\tfrac{1}{2}|\mathcal{C}_1^{-\frac{1}{2}}\mathbf{w}_1|^2\big).$$

Since

$$\begin{aligned}|\mathcal{C}_2^{-\frac{1}{2}}\mathbf{w}_2|^2-|\mathcal{C}_1^{-\frac{1}{2}}\mathbf{w}_1|^2&\overset{(3.53)}{=}\langle\hat{\mathcal{C}}_2^{-1}\mathbb{T}_{t-s}^{-1}\mathbf{w}_2,\mathbb{T}_{t-s}^{-1}\mathbf{w}_2\rangle-\langle\hat{\mathcal{C}}_1^{-1}\mathbb{T}_{t-s}^{-1}\mathbf{w}_1,\mathbb{T}_{t-s}^{-1}\mathbf{w}_1\rangle\\&\leqslant|\hat{\mathcal{C}}_1^{-1}||\mathbb{T}_{t-s}^{-1}(\mathbf{w}_2-\mathbf{w}_1)|\big(|\mathbb{T}_{t-s}^{-1}\mathbf{w}_2|+|\mathbb{T}_{t-s}^{-1}\mathbf{w}_1|\big)+|\hat{\mathcal{C}}_2^{-1}-\hat{\mathcal{C}}_1^{-1}||\mathbb{T}_{t-s}^{-1}\mathbf{w}_1|^2,\end{aligned}$$

by (3.18) and (3.54), we get

$$|\mathcal{C}_2^{-\frac{1}{2}}\mathbf{w}_2|^2-|\mathcal{C}_1^{-\frac{1}{2}}\mathbf{w}_1|^2\lesssim(t-s)^{\frac{\gamma}{2}}\mathcal{A}^{2+\gamma}.$$

Therefore, we have (3.52). The proof is complete. □

3.6 Gradient bounds in the non-degenerate direction $\boldsymbol{x}_1$

In this Section we provide pointwise controls for the derivatives of the densities in $\boldsymbol{x}_1$ in the current smooth setting. Importantly the controls do not depend on the smoothing procedure, and the constants therein only depend on Θ_T.

We are ready to prove the following

Proposition 3.10. *Under* $(\mathbf{H}_\sigma^\gamma)$, $(\mathbf{H}_\mathbf{F}^\gamma)$ *and* (3.1), *for any* $T>0$ *and* $j=0,1$, *there exists* $\lambda_j,C_j\geqslant1$ *depending on* Θ_T,η *such that for any* $0\leqslant s<t\leqslant T$ *and* $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^{2d}$,

$$|\nabla_{x_1}^j p(s,\boldsymbol{x};t,\boldsymbol{y})|\leqslant C_j(t-s)^{-\frac{j}{2}}\hat{p}_{\lambda_j}(s,\boldsymbol{x};t,\boldsymbol{y}),\tag{3.55}$$

and for $\eta_0,\eta_1\in(0,1)$, *and any* $\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y}\in\mathbb{R}^{2d}$,

$$|\nabla_{x_1}^j(p(s,\boldsymbol{x};t,\boldsymbol{y})-p(s,\boldsymbol{x}';t,\boldsymbol{y}))|\leqslant C_j|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\eta_j}(t-s)^{-\frac{j+\eta_j}{2}}\big(\hat{p}_{\lambda_j}(s,\boldsymbol{x};t,\boldsymbol{y})+\hat{p}_{\lambda_j}(s,\boldsymbol{x}';t,\boldsymbol{y})\big).\tag{3.56}$$

If in addition F_1 *also satisfies* (1.23), *then* (3.55) *and* (3.56) *also hold for* $j=2$ *and* $\eta_2\in(0,\gamma)$.

Proof. We only prove the case $j=2$. In this case some time singularities appear in the integrals. A way to overcome such a problem is to exploit cancellation properties of the derivatives of the Gaussian kernels. In the following estimates, for simplicity, we use the same λ to denote possible different constants in different places.

(i) We first look at (3.55). Starting from (3.44), for $u=\frac{t+s}{2}$ and $(\tau,\boldsymbol{\xi})\in[s,t]\times\mathbb{R}^{2d}$, we write

$$\begin{aligned}(\nabla_{x_1}^2\widetilde{p}_1\otimes\mathcal{H})(s,\boldsymbol{x};t,\boldsymbol{y})&=\int_u^t\int_{\mathbb{R}^{2d}}\nabla_{x_1}^2\widetilde{p}_1(s,\boldsymbol{x};r,\boldsymbol{z})\mathcal{H}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r\\&+\int_s^u\int_{\mathbb{R}^{2d}}\nabla_{x_1}^2\big(\widetilde{p}_1-\widetilde{p}^{(\tau,\boldsymbol{\xi})}\big)(s,\boldsymbol{x};r,\boldsymbol{z})\mathcal{H}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r\\&+\int_s^u\int_{\mathbb{R}^{2d}}\nabla_{x_1}^2\widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};r,\boldsymbol{z})\left(\mathcal{H}(r,\boldsymbol{z};t,\boldsymbol{y})-\mathcal{H}(r,\boldsymbol{\theta}_{r,\tau}(\boldsymbol{\xi});t,\boldsymbol{y})\right)\mathrm{d}\boldsymbol{z}\mathrm{d}r\\&=:I_1+I_2^{(\tau,\boldsymbol{\xi})}+I_3^{(\tau,\boldsymbol{\xi})}.\end{aligned}$$

For I_1, by (3.41) and (3.45) we have

$$I_1 \lesssim \int_u^t (r-s)^{-1} \int_{\mathbb{R}^{2d}} \bar{p}(s,\boldsymbol{x};r,\boldsymbol{z})(t-r)^{\frac{\gamma}{2}-1}\bar{p}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r$$
$$\leqslant 2(t-s)^{-1}\left(\int_u^t (t-r)^{\frac{\gamma}{2}-1}\mathrm{d}r\right)\bar{p}(s,\boldsymbol{x};t,\boldsymbol{y}) \lesssim (t-s)^{\frac{\gamma}{2}-1}\bar{p}(s,\boldsymbol{x};t,\boldsymbol{y}).$$

For $I_2^{(\tau,\boldsymbol{\xi})}$, taking $(\tau,\boldsymbol{\xi})=(s,\boldsymbol{x})$, by (3.49), (3.45) and (3.23), we have

$$\begin{aligned}
I_2^{(s,\boldsymbol{x})} &\lesssim \int_s^u \int_{\mathbb{R}^{2d}} (r-s)^{\frac{\gamma}{2}-1}\hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})(t-r)^{\frac{\gamma}{2}-1}\bar{p}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r \\
&\lesssim \left(\int_s^u (r-s)^{\frac{\gamma}{2}-1}(t-r)^{\frac{\gamma}{2}-1}\mathrm{d}r\right)\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}) \\
&\lesssim (t-s)^{\gamma-1}\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}).
\end{aligned}$$

For $I_3^{(\tau,\boldsymbol{\xi})}$, with the former choice $(\tau,\boldsymbol{\xi})=(s,\boldsymbol{x})$, by (3.49), (3.46), (3.41) and (3.23), we have

$$\begin{aligned}
I_3^{(s,\boldsymbol{x})} &\lesssim \int_s^u \int_{\mathbb{R}^{2d}} (r-s)^{-1}\hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})|\boldsymbol{z}-\boldsymbol{\theta}_{r,s}(\boldsymbol{x})|_{\mathbf{d}}^{\frac{\gamma}{2}}(t-r)^{\frac{\gamma}{2}-1} \\
&\qquad \times \Big[\bar{p}(r,\boldsymbol{z};t,\boldsymbol{y})+\bar{p}(r,\boldsymbol{\theta}_{r,s}(\boldsymbol{x});t,\boldsymbol{y})\Big]\mathrm{d}\boldsymbol{z}\mathrm{d}r \\
&\lesssim \int_s^u \int_{\mathbb{R}^{2d}} (r-s)^{\frac{\gamma}{4}-1}\hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})(t-r)^{\frac{\gamma}{2}-1} \\
&\qquad \times \Big[\hat{p}_\lambda(r,\boldsymbol{z};t,\boldsymbol{y})+\hat{p}_\lambda(r,\boldsymbol{\theta}_{r,s}(\boldsymbol{x});t,\boldsymbol{y})\Big]\mathrm{d}\boldsymbol{z}\mathrm{d}r \\
&\lesssim (t-s)^{\frac{\gamma}{2}-1}\left[\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y})+\int_s^u (r-s)^{\frac{\gamma}{4}-1}\hat{p}_\lambda(r,\boldsymbol{\theta}_{r,s}(\boldsymbol{x});t,\boldsymbol{y})\mathrm{d}r\right] \\
&\lesssim (t-s)^{\frac{\gamma}{2}-1}\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}),
\end{aligned}$$

where in the last step we have used that

$$\hat{p}_\lambda(r,\boldsymbol{\theta}_{r,s}(\boldsymbol{x});t,\boldsymbol{y}) = g_\lambda\big(t-r,\boldsymbol{\theta}_{t,r}\circ\boldsymbol{\theta}_{r,s}(\boldsymbol{x})-\boldsymbol{y}\big) \overset{(3.21)}{=} g_\lambda\big(t-r,\boldsymbol{\theta}_{t,s}(\boldsymbol{x})-\boldsymbol{y}\big). \tag{3.57}$$

Finally, note that

$$|\nabla_{x_1}^2 p(s,\boldsymbol{x};t,\boldsymbol{y})| \lesssim (t-s)^{-1}\bar{p}(s,\boldsymbol{x};t,\boldsymbol{y}) + |(\nabla_{x_1}^2\widetilde{p}_1\otimes\mathcal{H})(s,\boldsymbol{x};t,\boldsymbol{y})|,$$

which combining the above calculations, yields (3.55) for $j=2$.

(ii) Next we look at (3.56). We can assume w.l.o.g. that $|\boldsymbol{x}-\boldsymbol{x}'|\leqslant(t-s)^{\frac{1}{2}}$ since otherwise the control readily follows from the previous one. Starting from (3.44), for $u=\frac{t+s}{2}$ we write

$$\begin{aligned}
\nabla_{x_1}^2 p(s,\boldsymbol{x};t,\boldsymbol{y}) - \nabla_{x_1}^2 p(s,\boldsymbol{x}';t,\boldsymbol{y}) &= \nabla_{x_1}^2\widetilde{p}_1(s,\boldsymbol{x};t,\boldsymbol{y}) - \nabla_{x_1}^2\widetilde{p}_1(s,\boldsymbol{x}';t,\boldsymbol{y}) \\
&+ \int_s^t \int_{\mathbb{R}^{2d}} (\nabla_{x_1}^2(\widetilde{p}_1(s,\boldsymbol{x};r,\boldsymbol{z})-\nabla_{x_1}^2\widetilde{p}_1(s,\boldsymbol{x}';r,\boldsymbol{z}))\mathcal{H}(r,z;t,\boldsymbol{y})\mathrm{d}z\mathrm{d}r \\
&=: I_1 + \int_u^t \mathcal{J}(r,\boldsymbol{x},\boldsymbol{x}';t,\boldsymbol{y})\mathrm{d}r + \int_s^u \mathcal{J}(r,\boldsymbol{x},\boldsymbol{x}';t,\boldsymbol{y})\mathrm{d}r =: I_1+I_2+I_3.
\end{aligned}$$

For I_1, by (3.42) we have

$$I_1 \lesssim |\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\gamma}(t-s)^{-1-\frac{\gamma}{2}}\big(\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y})+\hat{p}_\lambda(s,\boldsymbol{x}';t,\boldsymbol{y})\big).$$

For I_2, by (3.42) and (3.45), we have

$$\begin{aligned} I_2 &\lesssim |\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\gamma}\int_u^t (r-s)^{-1-\frac{\gamma}{2}}(t-r)^{\frac{\gamma}{2}-1}\int_{\mathbb{R}^{2d}}\big(\bar{p}(s,\boldsymbol{x};r,\boldsymbol{z})+\bar{p}(s,\boldsymbol{x}';r,\boldsymbol{z})\big)\bar{p}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}z\mathrm{d}r \\ &\lesssim |\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\gamma}(t-s)^{-1}\big(\bar{p}(s,\boldsymbol{x};t,\boldsymbol{y})+\bar{p}(s,\boldsymbol{x}';t,\boldsymbol{y})\big). \end{aligned}$$

Next we treat the hard term I_3. Fix $r\in[s,u]$. We handle $\mathscr{J}$ according to the current *diagonal/off-diagonal* regime of $|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}$ w.r.t. to the current integration time and consider two cases:

$$\text{Case (I): } |\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}} > (r-s)^{\frac{1}{2}};\ \text{Case (II): } |\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}} \leqslant (r-s)^{\frac{1}{2}}.$$

In Case (I), for any $(\tau,\boldsymbol{\xi}),(\tau,\boldsymbol{\xi}')\in[s,t]\times\mathbb{R}^{2d}$, we write

$$\begin{aligned} \mathscr{J} &= \int_{\mathbb{R}^{2d}}\nabla_{x_1}^2\big(\widetilde{p}_1-\widetilde{p}^{(\tau,\boldsymbol{\xi})}\big)(s,\boldsymbol{x};r,\boldsymbol{z})\mathcal{H}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z} \\ &\quad+\int_{\mathbb{R}^{2d}}\nabla_{x_1}^2\widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};r,\boldsymbol{z})\big(\mathcal{H}(r,\boldsymbol{z};t,\boldsymbol{y})-\mathcal{H}(r,\boldsymbol{\theta}_{r,\tau}(\boldsymbol{\xi});t,\boldsymbol{y})\big)\mathrm{d}\boldsymbol{z} \\ &\quad-\int_{\mathbb{R}^{2d}}\nabla_{x_1}^2\big(\widetilde{p}_1-\widetilde{p}^{(\tau,\boldsymbol{\xi}')}\big)(s,\boldsymbol{x}';r,\boldsymbol{z})\mathcal{H}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z} \\ &\quad+\int_{\mathbb{R}^{2d}}\nabla_{x_1}^2\widetilde{p}^{(\tau,\boldsymbol{\xi}')}(s,\boldsymbol{x}';r,\boldsymbol{z})\big(\mathcal{H}(r,\boldsymbol{z};t,\boldsymbol{y})-\mathcal{H}(r,\boldsymbol{\theta}_{t,\tau}(\boldsymbol{\xi}');t,\boldsymbol{y})\big)\mathrm{d}\boldsymbol{z} \\ &=: \sum_{i=1,2}\Big(\mathscr{I}_i^{(\tau,\boldsymbol{\xi})}(r,\boldsymbol{x};t,\boldsymbol{y})+\mathscr{N}_i^{(\tau,\boldsymbol{\xi}')}(r,\boldsymbol{x}';t,\boldsymbol{y})\Big). \end{aligned}$$

By (3.49) and (3.45), we have

$$\begin{aligned} |\mathscr{I}_1^{(s,\boldsymbol{x})}(r,\boldsymbol{x};t,\boldsymbol{y})| &\lesssim (r-s)^{\frac{\gamma}{2}-1}(t-r)^{\frac{\gamma}{2}-1}\int_{\mathbb{R}^{2d}}\bar{p}(s,\boldsymbol{x};r,\boldsymbol{z})\bar{p}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}z \\ &= (r-s)^{\frac{\gamma}{2}-1}(t-r)^{\frac{\gamma}{2}-1}\bar{p}(s,\boldsymbol{x};t,\boldsymbol{y}). \end{aligned}$$

By (3.49) and (3.46), we have

$$\begin{aligned} |\mathscr{I}_2^{(s,\boldsymbol{x})}(r,\boldsymbol{x};t,\boldsymbol{y})| &\lesssim (r-s)^{-1}(t-r)^{\varepsilon-1}\int_{\mathbb{R}^{2d}}\hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})|\boldsymbol{z}-\boldsymbol{\theta}_{r,s}(\boldsymbol{x})|_{\mathbf{d}}^{\gamma-\varepsilon} \\ &\qquad\times\Big[\bar{p}(r,\boldsymbol{z};t,\boldsymbol{y})+\bar{p}(r,\boldsymbol{\theta}_{r,s}(\boldsymbol{x});t,\boldsymbol{y})\Big]\mathrm{d}z \\ &\overset{(3.22)}{\lesssim} (r-s)^{\frac{\gamma-\varepsilon}{2}-1}(t-r)^{\varepsilon-1}\int_{\mathbb{R}^{2d}}\hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z}) \\ &\qquad\times\Big[\hat{p}_\lambda(r,\boldsymbol{z};t,\boldsymbol{y})+\hat{p}_\lambda(r,\boldsymbol{\theta}_{r,s}(\boldsymbol{x});t,\boldsymbol{y})\Big]\mathrm{d}z \\ &\overset{(3.22),(3.57)}{\lesssim} (r-s)^{\frac{\gamma-\varepsilon}{2}-1}(t-r)^{\varepsilon-1}\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}). \end{aligned}$$

Similarly, we have

$$|\mathscr{N}_1^{(s,\boldsymbol{x}')}(r,\boldsymbol{x}';t,\boldsymbol{y})| \lesssim (r-s)^{\frac{\gamma}{2}-1}(t-r)^{\frac{\gamma}{2}-1}\bar{p}(s,\boldsymbol{x}';t,\boldsymbol{y}),$$
$$|\mathscr{N}_2^{(s,\boldsymbol{x}')}(r,\boldsymbol{x}';t,\boldsymbol{y})| \lesssim (r-s)^{\frac{\gamma-\varepsilon}{2}-1}(t-r)^{\varepsilon-1}\hat{p}_\lambda(s,\boldsymbol{x}';t,\boldsymbol{y}).$$

Combining the above estimates and taking ε small enough, we get

$$\int_s^u \mathbf{1}_{|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}>(r-s)^{1/2}}|\mathcal{J}(r,\boldsymbol{x},\boldsymbol{x}';t,\boldsymbol{y})|\mathrm{d}r \lesssim |\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\eta}(t-s)^{-1}\big(\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y})+\hat{p}_\lambda(s,\boldsymbol{x}';t,\boldsymbol{y})\big),$$

for all $\eta\in(0,\gamma)$.

In Case (II), for any $(\tau,\bar{\boldsymbol{\xi}})\in[s,t]\times\mathbb{R}^{2d}$, we write

$$\mathcal{J}=\int_0^1\Bigg[\int_{\mathbb{R}^{2d}}(\boldsymbol{x}-\boldsymbol{x}')\cdot\nabla_{\boldsymbol{x}}\nabla^2_{x_1}\big(\widetilde{p}_1-\widetilde{p}^{(\tau,\bar{\boldsymbol{\xi}})}\big)(s,\boldsymbol{x}'+\rho(\boldsymbol{x}-\boldsymbol{x}');r,\boldsymbol{z})\mathcal{H}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}$$
$$+\int_{\mathbb{R}^{2d}}(\boldsymbol{x}-\boldsymbol{x}')\cdot\nabla_{\boldsymbol{x}}\nabla^2_{x_1}\widetilde{p}^{(\tau,\bar{\boldsymbol{\xi}})}(s,\boldsymbol{x}'+\rho(\boldsymbol{x}-\boldsymbol{x}');r,\boldsymbol{z})\left(\mathcal{H}(r,\boldsymbol{z};t,\boldsymbol{y})-\mathcal{H}(r,\boldsymbol{\theta}_{r,s}(\bar{\boldsymbol{\xi}});t,\boldsymbol{y})\right)\mathrm{d}\boldsymbol{z}\Bigg]\mathrm{d}\rho.$$

In the above bracket, taking $\bar{\boldsymbol{\xi}}=\bar{\boldsymbol{\xi}}_\rho:=\boldsymbol{x}'+\rho(\boldsymbol{x}-\boldsymbol{x}')$ and by (3.49) and (3.46), we have

$$\begin{aligned}|\mathcal{J}|\lesssim&\int_0^1\Bigg[\sum_{i=1,2}\int_{\mathbb{R}^{2d}}|(\boldsymbol{x}-\boldsymbol{x}')_i|(r-s)^{\frac{\gamma}{2}-1-\frac{2i-1}{2}}\hat{p}_\lambda(s,\bar{\boldsymbol{\xi}}_\rho;r,\boldsymbol{z})(t-r)^{\frac{\gamma}{2}-1}\bar{p}(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\\&+\sum_{i=1,2}\int_{\mathbb{R}^{2d}}|(\boldsymbol{x}-\boldsymbol{x}')_i|(r-s)^{-1-\frac{2i-1}{2}}p_\lambda(s,\bar{\boldsymbol{\xi}}_\rho;r,\boldsymbol{z})|z-\boldsymbol{\theta}_{r,s}(\bar{\boldsymbol{\xi}}_\rho)|_{\mathbf{d}}^{\gamma-\varepsilon}\\&\times(t-r)^{\frac{\varepsilon}{2}-1}\Big[\hat{p}_\lambda(r,\boldsymbol{z};t,\boldsymbol{y})+\hat{p}_{\lambda_1}(r,\boldsymbol{\theta}_{r,s}(\bar{\boldsymbol{\xi}}_\rho);t,\boldsymbol{y})\Big]\mathrm{d}\boldsymbol{z}\Bigg]\mathrm{d}\rho\\\lesssim&|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\eta}(r-s)^{\frac{\gamma-\eta}{2}-1}(t-s)^{\frac{\varepsilon}{2}-1}\int_0^1\hat{p}_\lambda(s,\bar{\boldsymbol{\xi}}_\rho;t,\boldsymbol{y})\mathrm{d}\rho\\\lesssim&|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\eta}(r-s)^{\frac{\gamma-\eta}{2}-1}(t-s)^{\frac{\varepsilon}{2}-1}\hat{p}_\lambda(s,\boldsymbol{x}';t,\boldsymbol{y}),\end{aligned}$$

where in the second step we have used $r\in[s,u]$, $|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}\leqslant(r-s)^{1/2}$, (3.23) and (3.57). In the last step we have used (2.10). Therefore,

$$\int_s^u \mathbf{1}_{|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}\leqslant(r-s)^{1/2}}|\mathcal{J}(r,\boldsymbol{x},\boldsymbol{x}';t,\boldsymbol{y})|\mathrm{d}r \lesssim |\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\eta}(t-s)^{-1}\big(\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y})+\hat{p}_\lambda(s,\boldsymbol{x}';t,\boldsymbol{y})\big).$$

The whole proof is thus complete. □

3.7 Gradient bound in degenerate direction $\boldsymbol{x}_2$

The aim of this section is to show the following a priori gradient estimate.

Proposition 3.11. *Under* $(\mathbf{H}^\gamma_\sigma)$, $(\mathbf{H}^\gamma_{\mathbf{F}})$, (1.22), (1.23) *and* (3.1), *for any $T>0$, there exist constants $\lambda,C\geqslant1$ depending on Θ_T and $\bar{\kappa}_0,\bar{\kappa}_1$ such that for any $0\leqslant s<t\leqslant T$ and $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^{2d}$,*

$$|\nabla_{x_2}p(s,\boldsymbol{x};t,\boldsymbol{y})|\leqslant C(t-s)^{-\frac{3}{2}}\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}). \tag{3.58}$$

Proof. For $0 \leqslant s \leqslant t \leqslant T$, $\lambda > 0$ and $f \in C_b^\infty(\mathbb{R}^{2d})$, we define

$$P_{s,t}f(\boldsymbol{x}) := \int_{\mathbb{R}^{2d}} p(s,\boldsymbol{x};t,\boldsymbol{z})f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}, \quad \hat{P}^\lambda_{s,t}f(\boldsymbol{x}) := \int_{\mathbb{R}^{2d}} \hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{z})f(\boldsymbol{z})\mathrm{d}\boldsymbol{z},$$

and for $(\tau,\boldsymbol{\xi}) \in [s,t]\times\mathbb{R}^{2d}$,

$$\widetilde{P}^{(\tau,\boldsymbol{\xi})}_{s,t}f(\boldsymbol{x}) := \int_{\mathbb{R}^{2d}} \widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};t,\boldsymbol{z})f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}. \tag{3.59}$$

Under (3.1), it is standard to derive that

$$P_{s,t}f,\ \hat{P}^\lambda_{s,t}f,\ \widetilde{P}^{(\tau,\boldsymbol{\xi})}_{s,t}f \in C_b^\infty(\mathbb{R}^{2d}).$$

Thus all the calculations below are rigorous. Now we set

$$u := (s+t)/2.$$

Let $f \in C_b^\infty(\mathbb{R}^{2d})$ be nonnegative. Noting that

$$\widetilde{P}^{(\tau,\boldsymbol{\xi})}_{s,t}f = \widetilde{P}^{(\tau,\boldsymbol{\xi})}_{s,u}\widetilde{P}^{(\tau,\boldsymbol{\xi})}_{u,t}f,$$

by (3.16), it is easy to see that for any $(\tau,\boldsymbol{\xi}) \in [s,t]\times\mathbb{R}^{2d}$,

$$P_{s,t}f(\boldsymbol{x}) = \widetilde{P}^{(\tau,\boldsymbol{\xi})}_{s,u}P_{u,t}f(\boldsymbol{x}) + \int_s^u \widetilde{P}^{(\tau,\boldsymbol{\xi})}_{s,r}(\mathcal{L}_r - \mathcal{L}^{(\tau,\boldsymbol{\xi})}_r)P_{r,t}f(\boldsymbol{x})\mathrm{d}r. \tag{3.60}$$

Hence,

$$\nabla_{x_2}P_{s,t}f(\boldsymbol{x}) = \nabla_{x_2}\widetilde{P}^{(\tau,\boldsymbol{\xi})}_{s,u}P_{u,t}f(\boldsymbol{x}) + \int_s^u \nabla_{x_2}\widetilde{P}^{(\tau,\boldsymbol{\xi})}_{s,r}(\mathcal{L}_r - \mathcal{L}^{(\tau,\boldsymbol{\xi})}_r)P_{r,t}f(\boldsymbol{x})\mathrm{d}r. \tag{3.61}$$

Below we shall take $(\tau,\boldsymbol{\xi}) = (s,\boldsymbol{x})$. Observe that by (3.10) and (3.6),

$$\nabla_{x_2}\widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};r,\boldsymbol{z})\big|_{(\tau,\boldsymbol{\xi})=(s,\boldsymbol{x})} = -\nabla_{z_2}\widetilde{p}^{(\tau,\boldsymbol{\xi})}(s,\boldsymbol{x};r,\boldsymbol{z})\big|_{(\tau,\boldsymbol{\xi})=(s,\boldsymbol{x})} = -\nabla_{z_2}\widetilde{p}_0(s,\boldsymbol{x};r,\boldsymbol{z}).$$

In particular,

$$\nabla_{x_2}\widetilde{P}^{(\tau,\boldsymbol{\xi})}_{s,r}f(\boldsymbol{x})|_{(\tau,\boldsymbol{\xi})=(s,\boldsymbol{x})} = -\int_{\mathbb{R}^{2d}} \nabla_{z_2}\widetilde{p}_0(s,\boldsymbol{x};r,\boldsymbol{z})f(\boldsymbol{z})\mathrm{d}\boldsymbol{z} =: Q_{s,r}f(\boldsymbol{x}),$$

and for any bounded function $f(z_1)$ of the first variable,

$$Q_{s,r}f(\boldsymbol{x}) = -\int_{\mathbb{R}^{2d}} \nabla_{z_2}\widetilde{p}_0(s,\boldsymbol{x};r,\boldsymbol{z})f(z_1)\mathrm{d}\boldsymbol{z} \equiv 0. \tag{3.62}$$

Equation (3.62) is precisely what we call a partial cancellation property.

Moreover, by (3.15),

$$|\nabla_{z_2}\widetilde{p}_0(s,\boldsymbol{x};r,\boldsymbol{z})| \lesssim (r-s)^{-\frac{3}{2}}\hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z}) \Rightarrow |Q_{s,r}f(\boldsymbol{x})| \lesssim (r-s)^{-\frac{3}{2}}\hat{P}^\lambda_{s,r}f(\boldsymbol{x}). \tag{3.63}$$

Here and below, λ may vary from line to line but all the constants only depend on Θ_T and $\bar\kappa_0, \bar\kappa_1$. For notational convenience, we write

$$\mathcal{L}_r - \mathcal{L}_r^{(s,\boldsymbol{x})} = \mathscr{D}_r + \mathscr{B}_r + \mathscr{K}_r, \tag{3.64}$$

where

$$\mathscr{D}_r := \mathrm{tr}[(a(s,\boldsymbol{z}) - a(r,\boldsymbol{\theta}_{r,s}(\boldsymbol{x}))) \cdot \nabla_{z_1}^2]$$

and

$$\mathscr{B}_r := (F_1(r,\boldsymbol{z}) - F_1(r,\boldsymbol{\theta}_{r,s}(\boldsymbol{x}))) \cdot \nabla_{z_1}, \quad \mathscr{K}_r := \mathcal{T}_{F_2(r)}(\boldsymbol{z},\boldsymbol{\theta}_{r,s}(\boldsymbol{x})) \cdot \nabla_{z_2}.$$

Using the above notations, we can rewrite (3.61) as

$$\nabla_{x_2} P_{s,t} f(\boldsymbol{x}) = Q_{s,u} P_{u,t} f(\boldsymbol{x}) + \int_s^u Q_{s,r}(\mathscr{D}_r + \mathscr{B}_r + \mathscr{K}_r) P_{r,t} f(\boldsymbol{x}) \mathrm{d}r. \tag{3.65}$$

From (3.63), the upper bound for p and (3.23), we have

$$|Q_{s,u} P_{u,t} f(\boldsymbol{x})| \lesssim (u-s)^{-\frac{3}{2}} \hat{P}^\lambda_{s,u} P_{u,t} f(\boldsymbol{x}) \lesssim (t-s)^{-\frac{3}{2}} \hat{P}^\lambda_{s,t} f(\boldsymbol{x}).$$

To fully make up the time singularity in (3.63), we exploit the cancellation property (3.62). For $r \in [s,u]$, let $\Phi_r : \mathbb{R}^d \to [0,1]$ be a *smooth* cutoff function with

$$\Phi_r(z_2) = \begin{cases} 1, & |z_2 - \boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})| \leqslant (t-r)^{\frac{3}{2}}, \\ 0, & |z_2 - \boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})| \geqslant 2(t-r)^{\frac{3}{2}}, \end{cases} \quad |\nabla_{z_2} \Phi_r(z_2)| \leqslant 4(t-r)^{-\frac{3}{2}}. \tag{3.66}$$

First of all we look at the term containing $\mathscr{D}_r$ in (3.65) and write

$$\begin{aligned} Q_{s,r} \mathscr{D}_r P_{r,t} f(\boldsymbol{x}) &= -\int_{\mathbb{R}^{2d}} \nabla_{z_2} p_0(s,\boldsymbol{x};r,\boldsymbol{z})(1-\Phi_r(z_2)) \mathscr{D}_r P_{r,t} f(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} \\ &\quad + \int_{\mathbb{R}^{2d}} p_0(s,\boldsymbol{x};r,\boldsymbol{z}) \nabla_{z_2} \Phi_r(z_2) \mathscr{D}_r P_{r,t} f(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} \\ &\quad - \int_{\mathbb{R}^{2d}} \nabla_{z_2}(p_0(s,\boldsymbol{x};r,\boldsymbol{z}) \Phi_r(z_2)) \mathscr{D}_r P_{r,t} f(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} \\ &=: I_1 + I_2 - I_3. \end{aligned}$$

For I_1, by (3.63) and the upper bound estimate (3.55) of p, since f is nonnegative, we have

$$\begin{aligned} |I_1| &\lesssim (r-s)^{-\frac{3}{2}} \int_{\mathbb{R}^{2d}} \hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})(1-\Phi_r(z_2)) |\boldsymbol{z} - \boldsymbol{\theta}_{r,s}(\boldsymbol{x})|^\gamma_{\mathbf{d}} |\nabla^2_{z_1} P_{r,t} f(\boldsymbol{z})| \mathrm{d}\boldsymbol{z} \\ &\lesssim (r-s)^{\frac{\gamma-3}{2}} (t-r)^{-1} \int_{\mathbb{R}^{2d}} \hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z}) \mathbf{1}_{\{|z_2 - \boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})| \geqslant (t-r)^{3/2}\}} \hat{P}^\lambda_{r,t} f(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} \\ &\leqslant (r-s)^{\frac{\gamma-3}{2}} (t-r)^{-\frac{3}{2}} \int_{\mathbb{R}^{2d}} |z_2 - \boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})|^{\frac{1}{3}} \hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z}) \hat{P}^\lambda_{r,t} f(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} \\ &\lesssim (r-s)^{\frac{\gamma}{2}-1} (t-r)^{-\frac{3}{2}} \int_{\mathbb{R}^{2d}} \hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z}) \hat{P}^\lambda_{r,t} f(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} \\ &\lesssim (r-s)^{\frac{\gamma}{2}-1} (t-s)^{-\frac{3}{2}} \hat{P}^\lambda_{s,t} f(\boldsymbol{x}). \end{aligned}$$

Similarly, for I_2, we have

$$
\begin{aligned}
|I_2| &\lesssim \int_{\mathbb{R}^{2d}} \hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})|\nabla_{z_2}\Phi_r(z_2)|\,|\boldsymbol{z}-\boldsymbol{\theta}_{r,s}(\boldsymbol{x})|_{\mathbf{d}}^{\gamma}|\nabla^2_{z_1}P_{r,t}f(\boldsymbol{z})|\mathrm{d}\boldsymbol{z}\\
&\lesssim (t-r)^{-\frac{5}{2}}(r-s)^{\frac{\gamma}{2}}\int_{\mathbb{R}^{2d}} \hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})\hat{P}^\lambda_{r,t}f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\\
&\lesssim (t-s)^{\frac{\gamma-5}{2}}\hat{P}^\lambda_{s,t}f(\boldsymbol{x}).
\end{aligned}
$$

For I_3, similarly to the partial cancellation (3.62) we can write

$$
\begin{aligned}
I_3 = &\int_{\mathbb{R}^{2d}} \nabla_{z_2}(p_0(s,\boldsymbol{x};r,\boldsymbol{z})\Phi_r(z_2))\Big[\mathrm{tr}\left((a(r,\boldsymbol{z})-a(r,\boldsymbol{\theta}_{r,s}(\boldsymbol{x})))\cdot\nabla^2_{z_1}P_{r,t}f(\boldsymbol{z})\right)\\
&-\mathrm{tr}\left((a(r,z_1,\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x}))-a(r,\boldsymbol{\theta}_{r,s}(\boldsymbol{x})))\cdot\nabla^2_{z_1}P_{r,t}f(z_1,\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x}))\right)\Big]\mathrm{d}\boldsymbol{z}\\
= &\int_{\mathbb{R}^{2d}} \nabla_{z_2}(p_0(s,\boldsymbol{x};r,\boldsymbol{z})\Phi_r(z_2))\Big[\mathrm{tr}\left((a(r,\boldsymbol{z})-a(r,z_1,\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})))\cdot\nabla^2_{z_1}P_{r,t}f(\boldsymbol{z})\right)\\
&+\mathrm{tr}\big((a(r,z_1,\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x}))-a(r,\boldsymbol{\theta}_{r,s}(\boldsymbol{x})))\cdot\nabla^2_{z_1}(P_{r,t}f(\boldsymbol{z})-P_{r,t}f(z_1,\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})))\big)\Big]\mathrm{d}\boldsymbol{z}.
\end{aligned}
$$

Thus, by (3.63), (3.66), (3.56) and (1.22), for any $\eta\in(0,\gamma)$, we have

$$
\begin{aligned}
|I_3| &\lesssim \Big[(r-s)^{-\frac{3}{2}}+(t-s)^{-\frac{3}{2}}\Big]\int_{\mathbb{R}^{2d}} \hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})\mathbf{1}_{\{|z_2-\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})|\leqslant 2(t-r)^{3/2}\}}\\
&\quad\times\Big[|z_2-\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})|^{\frac{1+\gamma}{3}}|\nabla^2_{z_1}P_{r,t}f(\boldsymbol{z})|+|z_1-\boldsymbol{\theta}^1_{r,s}(\boldsymbol{x})|^{\alpha}|\nabla^2_{z_1}(P_{r,t}f(\boldsymbol{z})-P_{r,t}f(z_1,\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})))|\Big]\mathrm{d}\boldsymbol{z}\\
&\lesssim \Big[(r-s)^{-\frac{3}{2}}+(t-s)^{-\frac{3}{2}}\Big]\int_{\mathbb{R}^{2d}} \hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})\mathbf{1}_{\{|z_2-\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})|\leqslant 2(t-r)^{3/2}\}}\\
&\qquad\times\Big[|z_2-\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})|^{\frac{1+\gamma}{3}}(t-r)^{-1}\hat{P}^\lambda_{r,t}f(\boldsymbol{z})+|z_1-\boldsymbol{\theta}^1_{r,s}(\boldsymbol{x})|^{\alpha}|z_2-\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})|^{\frac{\eta}{3}}\\
&\qquad\quad\times(t-r)^{-1-\frac{\eta}{2}}(\hat{P}^\lambda_{r,t}f(\boldsymbol{z})+\hat{P}^\lambda_{r,t}f(z_1,\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})))\Big]\mathrm{d}\boldsymbol{z}\\
&\lesssim \Big[(r-s)^{-\frac{3}{2}}+(t-s)^{-\frac{3}{2}}\Big]\Big[(r-s)^{\frac{\gamma+1}{2}}(t-r)^{-1}+(r-s)^{\frac{\alpha+\eta}{2}}(t-r)^{-1-\frac{\eta}{2}}\Big]\\
&\qquad\times\int_{\mathbb{R}^{2d}} \hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})\hat{P}^\lambda_{r,t}f(\boldsymbol{z})\mathrm{d}z,
\end{aligned}
$$

where the last step is due to (3.22) and (2.10). Since $\alpha\in((1-\gamma)\vee\gamma,1]$, one can choose η close to γ so that for some $\varepsilon>0$,

$$
|I_3|\lesssim[(r-s)^{\varepsilon-1}+(t-s)^{\varepsilon-1}](t-s)^{-\frac{3}{2}}\hat{P}^\lambda_{s,t}f(\boldsymbol{x}).
$$

Next we treat the term containing $\mathscr{B}_r$, and similarly write

$$
\begin{aligned}
Q_{s,r}\mathscr{B}_rP_{r,t}f(\boldsymbol{x}) = &-\int_{\mathbb{R}^{2d}} \nabla_{z_2}p_0(s,\boldsymbol{x};r,\boldsymbol{z})(1-\Phi_r(z_2))\mathscr{B}_rP_{r,t}f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\\
&+\int_{\mathbb{R}^{2d}} p_0(s,\boldsymbol{x};r,\boldsymbol{z})\nabla_{z_2}\Phi_r(z_2)\mathscr{B}_rP_{r,t}f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\\
&-\int_{\mathbb{R}^{2d}} \nabla_{z_2}(p_0(s,\boldsymbol{x};r,\boldsymbol{z})\Phi_r(z_2))\mathscr{B}_rP_{r,t}f(\boldsymbol{z})\mathrm{d}\boldsymbol{z}
\end{aligned}
$$

$$=: J_1 + J_2 - J_3.$$

For J_1 and J_2, the analysis is similar to the one performed for I_1 and I_2. By (3.63) and the upper bound estimate (3.55) of p, we have

$$|J_1| + |J_2| \lesssim \Big[(r-s)^{\frac{\gamma}{2}-1}(t-r)^{-1} + (t-s)^{\frac{\gamma}{2}-2}\Big]\hat P^\lambda_{s,t} f(\boldsymbol{x}).$$

For J_3, as for I_3, similarly to (3.62) we can write

$$\begin{aligned} J_3 = \int_{\mathbb{R}^{2d}} &\nabla_{z_2}(p_0(s,\boldsymbol{x};r,\boldsymbol{z}) f_r(z_2))\Big[(F_1(r,\boldsymbol{z}) - F_1(r,z_1,\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})))\cdot\nabla_{z_1}P_{r,t}f(\boldsymbol{z}) \\ &+ (F_1(r,z_1,\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})) - F_1(r,\boldsymbol{\theta}_{r,s}(\boldsymbol{x})))\cdot\nabla_{z_1}(P_{r,t}f(\boldsymbol{z}) - P_{r,t}f(z_1,\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})))\Big]\mathrm{d}\boldsymbol{z}. \end{aligned}$$

Thus, by (3.63), (3.66), (3.56) and (1.23), for any $\eta \in (0,1)$, we have

$$\begin{aligned} |J_3| \lesssim{}& \Big[(r-s)^{-\frac{3}{2}} + (t-s)^{-\frac{3}{2}}\Big]\int_{\mathbb{R}^{2d}} \hat p_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})\mathbf{1}_{\{|z_2-\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})|\leqslant 2(t-r)^{3/2}\}} \\ &\times\Big[|z_2-\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})|^{\frac{1+\gamma}{3}}|\nabla_{z_1}P_{r,t}f(\boldsymbol{z})| + |z_1-\boldsymbol{\theta}^1_{r,s}(\boldsymbol{x})|^\gamma|\nabla_{z_1}(P_{r,t}f(\boldsymbol{z}) - P_{r,t}f(z_1,\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})))|\Big]\mathrm{d}\boldsymbol{z} \\ \lesssim{}& \Big[(r-s)^{-\frac{3}{2}} + (t-s)^{-\frac{3}{2}}\Big]\int_{\mathbb{R}^{2d}} \hat p_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})\mathbf{1}_{\{|z_2-\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})|\leqslant 2(t-r)^{3/2}\}} \\ &\times\Big[|z_2-\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})|^{\frac{1+\gamma}{3}}(t-r)^{-\frac{1}{2}}\hat P^\lambda_{r,t}f(\boldsymbol{z}) + |z_1-\boldsymbol{\theta}^1_{r,s}(\boldsymbol{x})|^\gamma|z_2-\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})|^{\frac{\eta}{3}} \\ &\times (t-r)^{-\frac{1}{2}-\frac{\eta}{2}}(\hat P^\lambda_{r,t}f(\boldsymbol{z}) + \hat P^\lambda_{r,t}f(z_1,\boldsymbol{\theta}^2_{r,s}(\boldsymbol{x})))\Big]\mathrm{d}\boldsymbol{z} \\ \lesssim{}& \Big[(r-s)^{-\frac{3}{2}} + (t-s)^{-\frac{3}{2}}\Big]\Big[(r-s)^{\frac{\gamma+1}{2}}(t-r)^{-\frac{1}{2}} + (r-s)^{\frac{\gamma+\eta}{2}}(t-r)^{-\frac{1}{2}-\frac{\eta}{2}}\Big] \\ &\times\int_{\mathbb{R}^{2d}} \hat p_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})\hat P^\lambda_{r,t}f(\boldsymbol{z})\mathrm{d}z, \end{aligned}$$

In particular, one can choose η close to 1 so that for some $\varepsilon > 0$,

$$|J_3| \lesssim [(r-s)^{\varepsilon-1} + (t-s)^{\varepsilon-1}](t-s)^{-1}\hat P^\lambda_{s,t}f(\boldsymbol{x}).$$

The term $Q_{s,r}\mathscr{B}_rP_{r,t}f(\boldsymbol{x})$ is therefore not critical in terms of the time singularities, since it appears to be less singular than $Q_{s,r}\mathscr{D}_rP_{r,t}f(\boldsymbol{x})$. However, the additional regularity assumption (1.23) is really needed to derive an integrable singularity in the variable r.

On the other hand, we have

$$\begin{aligned} |Q_{s,r}\mathscr{K}_rP_{r,t}f(\boldsymbol{x})| &\lesssim (r-s)^{-\frac{3}{2}}\int_{\mathbb{R}^{2d}} \hat p_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})|\boldsymbol{z}-\boldsymbol{\theta}_{r,s}(\boldsymbol{x})|^{1+\gamma}_{\mathbf{d}}|\nabla_{z_2}P_{r,t}f(\boldsymbol{z})|\mathrm{d}\boldsymbol{z} \\ &\lesssim (r-s)^{\frac{\gamma}{2}-1}\int_{\mathbb{R}^{2d}} \hat p_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})|\nabla_{z_2}P_{r,t}f(\boldsymbol{z})|\mathrm{d}\boldsymbol{z} \\ &= (r-s)^{\frac{\gamma}{2}-1}\hat P^\lambda_{s,r}|\nabla_{z_2}P_{r,t}f|(\boldsymbol{x}). \end{aligned}$$

Gathering together all the previous controls, we eventually get

$$|\nabla_{x_2}P_{s,t}f(\boldsymbol{x})| \lesssim (t-s)^{-\frac{3}{2}}\hat P^\lambda_{s,t}f(\boldsymbol{x}) + \int_s^u (r-s)^{\frac{\gamma}{2}-1}\hat P^\lambda_{s,r}|\nabla_{z_2}P_{r,t}f|(\boldsymbol{x})\mathrm{d}r$$

$$\lesssim (t-s)^{-\frac{3}{2}}\bar{P}^{\delta}_{s,t}f(\boldsymbol{x}) + \int_s^u (r-s)^{\frac{\gamma}{2}-1}\bar{P}^{\delta}_{s,r}|\nabla_{z_2}P_{r,t}f|(\boldsymbol{x})\mathrm{d}r, \tag{3.67}$$

where δ is chosen as in (3.40). Fix $s_0 < t$. For any $s \in (s_0, t]$, define

$$\Phi^f_{s_0,t}(s) := (t-s)^{\frac{3}{2}}\bar{P}^{\delta}_{s_0,s}|\nabla_{x_2}P_{s,t}f|(\boldsymbol{x}).$$

Using $\bar{P}^{\delta}_{s_0,s}$ act on both sides of (3.67) and by the Chapman-Kolmogorov equation, we obtain

$$\begin{aligned}
\Phi^f_{s_0,t}(s) &\lesssim \bar{P}^{\delta}_{s_0,t}f(\boldsymbol{x}) + (t-s)^{\frac{3}{2}}\int_s^u (r-s)^{\frac{\gamma}{2}-1}\bar{P}^{\delta}_{s_0,r}|\nabla_{z_2}P_{r,t}f|(\boldsymbol{x})\mathrm{d}r \\
&\leqslant \bar{P}^{\delta}_{s_0,t}f(\boldsymbol{x}) + (t-s)^{\frac{3}{2}}\int_s^u (r-s)^{\frac{\gamma}{2}-1}(t-r)^{-\frac{3}{2}}\Phi^f_{s_0,t}(r)\mathrm{d}r \\
&\leqslant \bar{P}^{\delta}_{s_0,t}f(\boldsymbol{x}) + 2^{\frac{3}{2}}\int_s^u (r-s)^{\frac{\gamma}{2}-1}\Phi^f_{s_0,t}(r)\mathrm{d}r \\
&\leqslant \bar{P}^{\delta}_{s_0,t}f(\boldsymbol{x}) + 2^{\frac{3}{2}}\int_s^t (r-s)^{\frac{\gamma}{2}-1}\Phi^f_{s_0,t}(r)\mathrm{d}r.
\end{aligned}$$

By the Volterra type Gronwall inequality we obtain

$$\Phi^f_{s_0,t}(s) \lesssim \bar{P}^{\delta}_{s_0,t}f(\boldsymbol{x}) \implies \bar{P}^{\delta}_{s_0,s}|\nabla_{x_2}P_{s,t}f|(\boldsymbol{x}) \lesssim (t-s)^{-\frac{3}{2}}\bar{P}^{\delta}_{s_0,t}f(\boldsymbol{x}). \tag{3.68}$$

Letting $s_0 \uparrow s$, we get

$$|\nabla_{x_2}P_{s,t}f|(\boldsymbol{x}) \lesssim (t-s)^{-\frac{3}{2}}\bar{P}^{\delta}_{s,t}f(\boldsymbol{x}).$$

Finally, for fixed $t' > t$ and $\boldsymbol{y} \in \mathbb{R}^{2d}$, we let $f(\boldsymbol{x}) := p(t,\boldsymbol{x};t',\boldsymbol{y}) \in C^\infty_b(\mathbb{R}^{2d})$, then by the Chapman-Kolmogorov equation and (3.23), we obtain

$$|\nabla_{x_2}p(s,\boldsymbol{x};t',\boldsymbol{y})| \lesssim_{C_3} (t-s)^{-\frac{3}{2}}\hat{p}_\lambda(s,\boldsymbol{x};t',\boldsymbol{y}).$$

This then readily gives estimate (3.58). The proof is complete. □

Proposition 3.12. *Under* $(\mathbf{H}^\gamma_\sigma)$, $(\mathbf{H}^\gamma_{\mathbf{F}})$, (1.22), (1.23) *and* (3.1), *for any* $T > 0$, *there exist constants* $\lambda, C \geqslant 1$ *and* $\eta \in (0, (\alpha-\gamma)\wedge(\alpha+\gamma-1))$, *depending on* Θ_T *and* $\bar\kappa_0, \bar\kappa_1$ *such that for any* $0 \leqslant s < t \leqslant T$ *and* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2d}$,

$$|\nabla_{x_2}(p(s,\boldsymbol{x};t,\boldsymbol{y}) - p(s,\boldsymbol{x}';t,\boldsymbol{y}))| \leqslant C|\boldsymbol{x}-\boldsymbol{x}'|^\eta_{\mathbf{d}}(t-s)^{-\frac{3+\eta}{2}}\big(\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}) + \hat{p}_\lambda(s,\boldsymbol{x}';t,\boldsymbol{y})\big). \tag{3.69}$$

Proof. If $|\boldsymbol{x}-\boldsymbol{x}'|^2_{\mathbf{d}} > t-s$, then by (3.58) we clearly have

$$|\nabla_{x_2}(p(s,\boldsymbol{x};t,\boldsymbol{y}) - p(s,\boldsymbol{x}';t,\boldsymbol{y}))| \lesssim (t-s)^{-\frac{3}{2}}(\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}) + \hat{p}_\lambda(s,\boldsymbol{x}';t,\boldsymbol{y})) \lesssim \text{r.h.s. of (3.69)}.$$

Next we restrict to the *global diagonal case* $|\boldsymbol{x}-\boldsymbol{x}'|^2_{\mathbf{d}} \leqslant t-s$. We first write a localized version of the Duhamel formula. For any freezing couple $(\bar\tau, \bar{\boldsymbol{\xi}})$ we have taking formally $f = \delta_{\boldsymbol{y}}$ in (3.60)

$$p(s,\boldsymbol{x};t,\boldsymbol{y}) = \tilde{P}^{(\bar\tau,\bar{\boldsymbol{\xi}})}_{s,u}p(u,\cdot;t,\boldsymbol{y})(\boldsymbol{x}) + \int_s^u\int_{\mathbb{R}^{2d}}\tilde{p}^{(\bar\tau,\bar{\boldsymbol{\xi}})}(s,\boldsymbol{x};r,\boldsymbol{z})\bar{\mathscr{L}}^{(\bar\tau,\bar{\boldsymbol{\xi}})}p(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r, \tag{3.70}$$

where $\widetilde{P}^{(\bar\tau,\bar{\boldsymbol{\xi}})}_{t,u}f$ is as in (3.59) and we denoted $\bar{\mathscr{L}}^{(\bar\tau,\bar{\boldsymbol{\xi}})} = \mathscr{L} - \mathscr{L}^{(\bar\tau,\bar{\boldsymbol{\xi}})}$. Let us now differentiate w.r.t u: we obtain

$$0 = \partial_u[\widetilde{P}^{(\bar\tau,\bar{\boldsymbol{\xi}})}_{s,u}p(u,\cdot;t,\boldsymbol{y})(\boldsymbol{x})] + \int_{\mathbb{R}^{2d}}\tilde{p}^{(\bar\tau,\bar{\boldsymbol{\xi}})}(s,\boldsymbol{x};u,\boldsymbol{z})\bar{\mathscr{L}}^{(\bar\tau,\bar{\boldsymbol{\xi}})}p(u,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}. \tag{3.71}$$

We integrate the previous equation taking:

1. $(\bar{\tau}, \bar{\boldsymbol{\xi}}) = (\tau_0, \boldsymbol{\xi}_0)$ between s and $s_1 = \frac{t+s}{2}$;
2. $(\bar{\tau}, \bar{\boldsymbol{\xi}}) = (\tau_1, \boldsymbol{\xi}_1)$ between s_1 and t;

Then we get

$$0 = \tilde{P}^{(\tau_0,\boldsymbol{\xi}_0)}_{s,s_1} p(s_1, \cdot; t, \boldsymbol{y})(\boldsymbol{x}) - p(s, \boldsymbol{x}; t, \boldsymbol{y}) + \int_s^{s_1} \int_{\mathbb{R}^{2d}} \tilde{p}^{(\tau_0,\boldsymbol{\xi}_0)}(s, \boldsymbol{x}; r, \boldsymbol{z}) \bar{\mathcal{L}}^{(\tau_0,\boldsymbol{\xi}_0)} p(r, \boldsymbol{z}; t, \boldsymbol{y}) \mathrm{d}\boldsymbol{z}\mathrm{d}r,$$
$$0 = \tilde{p}^{(\tau_1,\boldsymbol{\xi}_1)}(s, \boldsymbol{x}; t, \boldsymbol{y}) - \tilde{P}^{(\tau_1,\boldsymbol{\xi}_1)}_{s,s_1} p(s_1, \cdot; t, \boldsymbol{y})(\boldsymbol{x}) + \int_{s_1}^{t} \int_{\mathbb{R}^{2d}} \tilde{p}^{(\tau_1,\boldsymbol{\xi}_1)}(s, \boldsymbol{x}; r, \boldsymbol{z}) \bar{\mathcal{L}}^{(\tau_1,\boldsymbol{\xi}_1)} p(r, \boldsymbol{z}; t, \boldsymbol{y}) \mathrm{d}\boldsymbol{z}\mathrm{d}r.$$

Next we use expansion (3.71) for $p(t, \boldsymbol{x}'; s, \boldsymbol{y})$ and integrate the equation taking:

1. $(\bar{\tau}, \bar{\boldsymbol{\xi}}) = (\tau_0, \boldsymbol{\xi}_0')$ between s and $s_0 = t + c|\boldsymbol{x} - \boldsymbol{x}'|^2_{\mathbf{d}}$;
2. $(\bar{\tau}, \bar{\boldsymbol{\xi}}) = (\tau_0, \boldsymbol{\xi}_0)$ between s_0 and s_1;
3. $(\bar{\tau}, \bar{\boldsymbol{\xi}}) = (\tau_1, \boldsymbol{\xi}_1)$ between s_1 and t;

Then we get

$$0 = \tilde{P}^{(\tau_0,\boldsymbol{\xi}_0')}_{s,s_0} p(s_0, \cdot; t, \boldsymbol{y})(\boldsymbol{x}') - p(s, \boldsymbol{x}'; t, \boldsymbol{y}) + \int_s^{s_0} \int_{\mathbb{R}^{2d}} \tilde{p}^{(\tau_0,\boldsymbol{\xi}_0')}(s, \boldsymbol{x}'; r, \boldsymbol{z}) \bar{\mathcal{L}}^{(\tau_0,\boldsymbol{\xi}_0')} p(r, \boldsymbol{z}; s, \boldsymbol{y}) \mathrm{d}\boldsymbol{z}\mathrm{d}r,$$
$$0 = \tilde{P}^{(\tau_0,\boldsymbol{\xi}_0)}_{s,s_1} p(s_1, \cdot; t, \boldsymbol{y})(\boldsymbol{x}') - \tilde{P}^{(\tau_0,\boldsymbol{\xi}_0)}_{s,s_0} p(s_0, \cdot; t, \boldsymbol{y})(\boldsymbol{x}') + \int_{s_0}^{s_1} \int_{\mathbb{R}^{2d}} \tilde{p}^{(\tau_0,\boldsymbol{\xi}_0)}(s, \boldsymbol{x}'; r, \boldsymbol{z}) \bar{\mathcal{L}}^{(\tau_0,\boldsymbol{\xi}_0)} p(r, \boldsymbol{z}; t, \boldsymbol{y}) \mathrm{d}\boldsymbol{z}\mathrm{d}r.$$
$$0 = \tilde{p}^{(\tau_1,\boldsymbol{\xi}_1)}(s, \boldsymbol{x}'; t, \boldsymbol{y}) - \tilde{P}^{(\tau_1,\boldsymbol{\xi}_1)}_{s,s_1} p(s_1, \cdot; t, \boldsymbol{y})(\boldsymbol{x}') + \int_{s_1}^{t} \int_{\mathbb{R}^{2d}} \tilde{p}^{(\tau_1,\boldsymbol{\xi}_1)}(s, \boldsymbol{x}'; r, \boldsymbol{z}) \bar{\mathcal{L}}^{(\tau_1,\boldsymbol{\xi}_1)} p(r, \boldsymbol{z}; t, \boldsymbol{y}) \mathrm{d}\boldsymbol{z}\mathrm{d}r.$$

Notice that it suffices to take $0 < c < \frac{1}{2}$ to ensure $s_0 < s_1$. We then have:

$$\begin{aligned}
&p(s, \boldsymbol{x}; t, \boldsymbol{y}) - p(s, \boldsymbol{x}'; t, \boldsymbol{y}) \\
&\quad = \tilde{p}^{(\tau_1,\boldsymbol{\xi}_1)}(s, \boldsymbol{x}; t, \boldsymbol{y}) - \tilde{p}^{(\tau_1,\boldsymbol{\xi}_1)}(s, \boldsymbol{x}'; t, \boldsymbol{y}) \\
&\qquad + \Delta\tilde{P}^{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0')}(s_0, t; \boldsymbol{x}', \boldsymbol{x}', \boldsymbol{y}) + \Delta\tilde{P}^{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0)}(s_1, t; \boldsymbol{x}, \boldsymbol{x}', \boldsymbol{y}) + \Delta\tilde{P}^{(\tau_1,\boldsymbol{\xi}_1,\boldsymbol{\xi}_1)}(s_1, t; \boldsymbol{x}', \boldsymbol{x}, \boldsymbol{y}) \\
&\qquad + \Delta^{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0')}_{\mathrm{off-diag}}(s, t; \boldsymbol{x}, \boldsymbol{x}', \boldsymbol{y}) + \Delta^{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0)}_{\mathrm{diag}}(s, t; \boldsymbol{x}, \boldsymbol{x}', \boldsymbol{y})
\end{aligned}$$

where

$$\begin{aligned}
\Delta\tilde{P}^{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0')}(s_0, t; \boldsymbol{x}', \boldsymbol{x}', \boldsymbol{y}) &= -\tilde{P}^{(\tau_0,\boldsymbol{\xi}_0')}_{s,s_0} p(s_0, \cdot; t, \boldsymbol{y})(\boldsymbol{x}') + \tilde{P}^{(\tau_0,\boldsymbol{\xi}_0)}_{s,s_0} p(s_0, \cdot; t, \boldsymbol{y})(\boldsymbol{x}'), \\
\Delta\tilde{P}^{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0)}(s_1, t; \boldsymbol{x}, \boldsymbol{x}', \boldsymbol{y}) &= \tilde{P}^{(\tau_0,\boldsymbol{\xi}_0)}_{s,s_1} p(s_1, \cdot; t, \boldsymbol{y})(\boldsymbol{x}) - \tilde{P}^{(\tau_0,\boldsymbol{\xi}_0)}_{s,s_1} p(s_1, \cdot; t, \boldsymbol{y})(\boldsymbol{x}'), \\
\Delta\tilde{P}^{(\tau_1,\boldsymbol{\xi}_1,\boldsymbol{\xi}_1)}(s_1, t; \boldsymbol{x}', \boldsymbol{x}, \boldsymbol{y}) &= \tilde{P}^{(\tau_1,\boldsymbol{\xi}_1)}_{s,s_1} p(s_1, \cdot; t, \boldsymbol{y})(\boldsymbol{x}') - \tilde{P}^{(\tau_1,\boldsymbol{\xi}_1)}_{s,s_1} p(s_1, \cdot; t, \boldsymbol{y})(\boldsymbol{x}),
\end{aligned}$$

and

$$\begin{aligned}
\Delta^{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0')}_{\mathrm{off-diag}}(s, t; \boldsymbol{x}, \boldsymbol{x}', \boldsymbol{y}) = &\int_s^{s_0} \int_{\mathbb{R}^{2d}} \tilde{p}^{(\tau_0,\boldsymbol{\xi}_0)}(s, \boldsymbol{x}; r, \boldsymbol{z}) \bar{\mathcal{L}}^{(\tau_0,\boldsymbol{\xi}_0)} p(r, \boldsymbol{z}; t, \boldsymbol{y}) \mathrm{d}\boldsymbol{z}\mathrm{d}r \\
&- \int_s^{s_0} \int_{\mathbb{R}^{2d}} \tilde{p}^{(\tau_0,\boldsymbol{\xi}_0')}(s, \boldsymbol{x}'; r, \boldsymbol{z}) \bar{\mathcal{L}}^{(\tau_0,\boldsymbol{\xi}_0')} p(r, \boldsymbol{z}; t, \boldsymbol{y}) \mathrm{d}\boldsymbol{z}\mathrm{d}r,
\end{aligned}$$

$$
\begin{aligned}
\Delta_{\rm diag}^{(\tau_0,\boldsymbol{\xi}_0,\tau_1,\boldsymbol{\xi}_1)}(s,t;\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y}) &= \int_{s_0}^{s_1}\int_{\mathbb{R}^{2d}}(\tilde{p}^{(\tau_0,\boldsymbol{\xi}_0)}(s,\boldsymbol{x};r,\boldsymbol{z})-\tilde{p}^{(\tau_0,\boldsymbol{\xi}_0)}(s,\boldsymbol{x}';r,\boldsymbol{z}))\bar{\mathcal{L}}^{(\tau_0,\boldsymbol{\xi}_0)}p(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r,\\
&\quad+\int_{s_1}^{t}\int_{\mathbb{R}^{2d}}(\tilde{p}^{(\tau_1,\boldsymbol{\xi}_1)}(s,\boldsymbol{x};r,\boldsymbol{z})-\tilde{p}^{(\tau_1,\boldsymbol{\xi}_1)}(s,\boldsymbol{x}';r,\boldsymbol{z}))\bar{\mathcal{L}}^{(\tau_1,\boldsymbol{\xi}_1)}p(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r\\
&=: \Delta_{\rm diag,1}^{(\tau_0,\boldsymbol{\xi}_0)}(s,t;\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y})+\Delta_{\rm diag,2}^{(\tau_1,\boldsymbol{\xi}_1)}(s,t;\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y}).
\end{aligned}
$$

Let us start with the term $\Delta_{\rm off-diag}^{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0')}(s,t;\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y})$. Differentiating with respect to x_2 and taking $(\tau,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0')=(s,\boldsymbol{x},\boldsymbol{x}')$ **after** differentiation, we have

$$
\begin{aligned}
&|\nabla_{x_2}\Delta_{\rm off-diag}^{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0')}(s,t;\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y})|_{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0')=(s,\boldsymbol{x},\boldsymbol{x}')}\\
&\quad\leqslant \left|\int_s^{s_0}\int_{\mathbb{R}^{2d}}\nabla_{\boldsymbol{x}_2}\tilde{p}^{(\tau_0,\boldsymbol{\xi}_0)}(s,\boldsymbol{x};r,\boldsymbol{z})\bar{\mathcal{L}}^{(\tau_0,\boldsymbol{\xi}_0)}p(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r\right|_{(\tau_0,\boldsymbol{\xi}_0)=(s,\boldsymbol{x})}\\
&\qquad+\left|\int_s^{s_0}\int_{\mathbb{R}^{2d}}\nabla_{\boldsymbol{x}_2}\tilde{p}^{(\tau_0,\boldsymbol{\xi}_0')}(s,\boldsymbol{x}';r,\boldsymbol{z})\bar{\mathcal{L}}^{(\tau_0,\boldsymbol{\xi}_0')}p(r,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\mathrm{d}r\right|_{(\tau_0,\boldsymbol{\xi}_0')=(s,\boldsymbol{x}')},
\end{aligned}
$$

where both the terms in the r.h.s. can be controlled separately as in the proof of Proposition 3.11. We derive:

$$
\begin{aligned}
&|\nabla_{x_2}\Delta_{\rm off-diag}^{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0')}(s,t;\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y})|_{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0')=(s,\boldsymbol{x},\boldsymbol{x}')}\\
&\quad\lesssim (\bar{p}(s,\boldsymbol{x};t,\boldsymbol{y})+\bar{p}(s,\boldsymbol{x}';t,\boldsymbol{y}))\int_s^{s+c|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^2}\left[(r-s)^{-1+\frac{\gamma}{2}}+(r-s)^{\frac{\alpha+\eta-3}{2}}\right](t-r)^{-\frac{3}{2}}\mathrm{d}r
\end{aligned}
$$

(since $\alpha>(\gamma\vee 1-\gamma)$, we can choose any $0<\epsilon<(\alpha-\gamma)\wedge(\alpha+\gamma-1)$ such that)

$$
\begin{aligned}
&\lesssim (\bar{p}(s,\boldsymbol{x};t,\boldsymbol{y})+\bar{p}(s,\boldsymbol{x}';t,\boldsymbol{y}))\int_s^{s+c|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^2}(r-s)^{-1+\frac{\epsilon}{2}}(t-r)^{-\frac{3}{2}}\mathrm{d}r\\
&\lesssim (t-r)^{-\frac{3}{2}}|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\epsilon}(\bar{p}(s,\boldsymbol{x};t,\boldsymbol{y})+\bar{p}(s,\boldsymbol{x}';t,\boldsymbol{y})).
\end{aligned}
$$

Let us now turn to $\Delta_{\rm diag}^{(\tau_0,\boldsymbol{\xi}_0,\tau_1,\boldsymbol{\xi}_1)}(s,t;\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y})$: similarly to (3.27), in the *local diagonal case* $|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^2\leqslant r-s$ we have, for any $\alpha\in[0,1]$,

$$
\begin{aligned}
&|\nabla_{x_2}\left(\widetilde{p}^{(\tau_0,\boldsymbol{\xi}_0)}(s,\boldsymbol{x};r,\boldsymbol{z})-\widetilde{p}^{(\tau_0,\boldsymbol{\xi}_0)}(s,\boldsymbol{x}';r,\boldsymbol{z})\right)|_{(\tau_0,\boldsymbol{\xi}_0)=(s,\boldsymbol{x})}\\
&\quad\leqslant\sum_{i=1}^{2}|x_i-x_i'|\sup_{\eta\in[0,1]}|\nabla_{x_i}\nabla_{x_2}\widetilde{p}^{(\tau_0,\boldsymbol{\xi}_0)}(s,\boldsymbol{x}+\eta(\boldsymbol{x}'-\boldsymbol{x});r,\boldsymbol{z})|_{(\tau_0,\boldsymbol{\xi}_0)=(s,\boldsymbol{x})}\\
&\quad\lesssim|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\alpha}(r-s)^{-\frac{3+\alpha}{2}}\hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z}).
\end{aligned}
$$

Therefore we have

$$
\begin{aligned}
&|\nabla_{x_2}\Delta_{\rm diag,1}^{(\tau_0,\boldsymbol{\xi}_0)}(s,t;\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y})|_{(\tau_0,\boldsymbol{\xi}_0)=(s,\boldsymbol{x})}\\
&\quad\lesssim|\boldsymbol{x}-\boldsymbol{x}'|_{\mathbf{d}}^{\eta}\int_{s_0}^{s_1}(r-s)^{-\frac{3+\eta}{2}}\int_{\mathbb{R}^{2d}}\hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})|\bar{\mathcal{L}}^{(s,\boldsymbol{x})}p(r,\boldsymbol{z};t,\boldsymbol{y})|\mathrm{d}\boldsymbol{z}\mathrm{d}r,
\end{aligned}
$$

which is controlled again, as in the proof of Proposition 3.11. We eventually derive, for any ϵ as above, and $\eta < \epsilon$

$$
\begin{aligned}
|\nabla_{x_2}\Delta^{(\tau_0,\boldsymbol{\xi}_0)}_{\text{diag},1}(s,t;\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y})|_{(\tau_0,\boldsymbol{\xi}_0)=(s,\boldsymbol{x})} &\lesssim |\boldsymbol{x}-\boldsymbol{x}'|^{\eta}_{\mathbf{d}}\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y})\int_{s_0}^{s_1}(r-s)^{-\frac{1+\eta-\epsilon}{2}}(t-r)^{-\frac{3}{2}}\mathrm{d}r\\
&\lesssim |\boldsymbol{x}-\boldsymbol{x}'|^{\eta}_{\mathbf{d}}(t-s)^{-\frac{3}{2}}\hat{p}_\lambda(s,\boldsymbol{x};t,\boldsymbol{y}).
\end{aligned}
$$

The control for $\Delta^{(\tau_1,\boldsymbol{\xi}_1)}_{\text{diag},2}(s,t;\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y})$ is more direct. Indeed, taking $(\tau_1,\boldsymbol{\xi}_1)=(t,\boldsymbol{y})$ we have, by (3.27),

$$
\begin{aligned}
&|\nabla_{x_2}\Delta^{(\tau_1,\boldsymbol{\xi}_1)}_{\text{diag},2}(s,t;\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y})|_{(\tau_1,\boldsymbol{\xi}_1)=(t,\boldsymbol{y})}\\
&\quad\lesssim \int_{s_1}^{t}\int_{\mathbb{R}^{2d}}\frac{|\boldsymbol{x}-\boldsymbol{x}'|^{\eta}_{\mathbf{d}}}{(r-s)^{\frac{3+\eta}{2}}}(\hat{p}_\lambda(s,\boldsymbol{x};r,\boldsymbol{z})+\hat{p}_\lambda(s,\boldsymbol{x}';r,\boldsymbol{z}))\frac{\hat{p}_\lambda(r,\boldsymbol{z};t,\boldsymbol{z})}{(t-r)^{1-\frac{\gamma}{2}}}\mathrm{d}\boldsymbol{z}\mathrm{d}r\\
&\quad\lesssim |\boldsymbol{x}-\boldsymbol{x}'|^{\eta}_{\mathbf{d}}(t-s)^{-\frac{3+\eta}{2}}(\hat{p}_{\lambda'}(s,\boldsymbol{x};t,\boldsymbol{y})+\hat{p}_{\lambda'}(s,\boldsymbol{x}';t,\boldsymbol{y}))
\end{aligned}
$$

where we used (3.23) in the last inequality.

It remains to check the terms $\Delta\tilde{P}$, which arise from the change of freezing couples. By (3.50), we have

$$
\begin{aligned}
&|\nabla_{x_2}\Delta\tilde{P}^{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0')}(s_0,t;\boldsymbol{x}',\boldsymbol{x}',\boldsymbol{y})|_{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0')=(s,\boldsymbol{x},\boldsymbol{x}')}\\
&\quad= \left|\nabla_{x_2}\int_{\mathbb{R}^{2d}}\Big(\tilde{p}^{(\tau_0,\boldsymbol{\xi}_0)}(s,\boldsymbol{x}';s_0,\boldsymbol{z})-\tilde{p}^{(\tau_0,\boldsymbol{\xi}_0')}(s,\boldsymbol{x}';s_0,\boldsymbol{z})\Big)p(s_0,\boldsymbol{z};t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}\right|_{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0')=(s,\boldsymbol{x},\boldsymbol{x}')}\\
&\quad\lesssim |\boldsymbol{x}-\boldsymbol{x}'|^{\gamma-\epsilon}\int_{\mathbb{R}^{2d}}(s_0-s)^{-\frac{3-\epsilon}{2}}\hat{p}_\lambda(s,\boldsymbol{x}';s_0,\boldsymbol{z})(p(s_0,\boldsymbol{z};t,\boldsymbol{y})-p(s_0,\boldsymbol{z}_1,\boldsymbol{\theta}^2_{s_0,s}(\boldsymbol{x}');t,\boldsymbol{y}))\mathrm{d}\boldsymbol{z}\\
&\quad\lesssim \frac{|\boldsymbol{x}-\boldsymbol{x}'|^{\gamma-\epsilon}}{(s_0-s)^{\frac{3-\epsilon}{2}}}\int_{\mathbb{R}^{2d}}\hat{p}_\lambda(s,\boldsymbol{x}';s_0,\boldsymbol{z})\frac{|\boldsymbol{z}_2-\boldsymbol{\theta}^2_{s_0,s}(\boldsymbol{x}')|}{(t-s_0)^{\frac{3}{2}}}\sup_{\rho\in[0,1]}\hat{p}_\lambda(s_0,\boldsymbol{z}_1,\boldsymbol{z}_2+\rho(\boldsymbol{\theta}^2_{s_0,s}(\boldsymbol{x}')-\boldsymbol{z}_2);t,\boldsymbol{y})\mathrm{d}\boldsymbol{z}
\end{aligned}
$$

(by (3.22), choosing $\epsilon<\gamma$, and since $2(t-s_0)>t-s$)

$$
\lesssim \frac{|\boldsymbol{x}-\boldsymbol{x}'|^{\gamma-\epsilon}}{(t-s)^{\frac{3}{2}}}\hat{p}_{\lambda'}(s,\boldsymbol{x}';t,\boldsymbol{y}).
$$

The remaining terms $\Delta\tilde{P}^{(\tau_0,\boldsymbol{\xi}_0,\boldsymbol{\xi}_0)}(s_1,t;\boldsymbol{x},\boldsymbol{x}',\boldsymbol{y})$ and $\Delta\tilde{P}^{(\tau_1,\boldsymbol{\xi}_1,\boldsymbol{\xi}_1)}(s_1,t;\boldsymbol{x}',\boldsymbol{x},\boldsymbol{y})$ are handled similarly, following (3.22) and Lemma 3.49. □

4 Proof of the main theorem

In this section we assume $(\mathbf{H}^{\gamma}_{\sigma})$ and $(\mathbf{H}^{\gamma}_{\mathbf{F}})$ for some $\gamma\in(0,1)$. For $\varepsilon\in(0,1)$, we define

$$
\mathbf{F}^{(\varepsilon)}(t,\boldsymbol{x}) := \mathbf{F}(t,\cdot)*\rho_\varepsilon(\boldsymbol{x}) = (F_1^{(\varepsilon)},F_2^{(\varepsilon)})(t,\boldsymbol{x})
$$

and

$$
\sigma^{(\varepsilon)}(s,\boldsymbol{x}) := \sigma(t,\cdot)*\rho_\varepsilon(\boldsymbol{x}).
$$

Let $\mathbf{X}^\varepsilon_{t,s}(\boldsymbol{x})$ be the solution of (1.3) corresponding to $(\mathbf{F}^{(\varepsilon)}, \sigma^{(\varepsilon)})$ and p_ε be the corresponding density. By Theorem 11.1.4 in [38] and Theorem 1 in [6] , under $(\mathbf{H}^\gamma_{\mathbf{F}})$ and $(\mathbf{H}^\gamma_\sigma)$, for any $f \in C^\infty_c(\mathbb{R}^{2d})$,

$$\lim_{\varepsilon\to 0} \mathbb{E} f(\mathbf{X}^\varepsilon_{t,s}(\boldsymbol{x})) = \mathbb{E} f(\mathbf{X}_{t,s}(\boldsymbol{x})). \tag{4.1}$$

Moreover, from Propositions 3.6, 3.7 and Section 3.3 we have the following uniform estimate: there exist constants $\lambda_0, C > 0$ depending only on Θ_T such that for all $\varepsilon \in (0, 1)$,

$$C^{-1} g_{\lambda_0^{-1}}\big(t-s, \boldsymbol{\theta}^{(\varepsilon)}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big) \leqslant p_\varepsilon(s, \boldsymbol{x}; t, \boldsymbol{y}) \leqslant C g_{\lambda_0}\big(t-s, \boldsymbol{\theta}^{(\varepsilon)}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big), \tag{4.2}$$

where $\boldsymbol{\theta}^{(\varepsilon)}_{t,s}(\boldsymbol{x})$ is the unique solution of the following ODE

$$\dot{\boldsymbol{\theta}}^{(\varepsilon)}_{t,s}(\boldsymbol{x}) = \mathbf{F}^{(\varepsilon)}(t, \boldsymbol{\theta}^{(\varepsilon)}_{t,s}(\boldsymbol{x})), \quad \boldsymbol{\theta}^{(\varepsilon)}_{s,s}(\boldsymbol{x}) = \boldsymbol{x}.$$

In order to take limits $\varepsilon \to 0$, we need the following important estimate.

Lemma 4.1. *For any $T > 0$, there exists a constant $C \geqslant 1$ only depending on Θ_T such that for all $0 \leqslant s < t \leqslant T$, $\boldsymbol{x} \in \mathbb{R}^{2d}$ and $\varepsilon \leqslant (t-s)^{3/2}$,*

$$|\mathbb{T}^{-1}_{t-s}\big(\boldsymbol{\theta}^{(\varepsilon)}_{t,s}(\boldsymbol{x}) - \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x})\big)| \leqslant C, \tag{4.3}$$

where $\widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x})$ is defined by ODE (1.15).

Proof. For simplicity of notations, we assume $s = 0$ and write for $\boldsymbol{x} \in \mathbb{R}^{2d}$ and $t \geqslant 0$,

$$\ell_i(t) := \big|\big(\boldsymbol{\theta}^{(\varepsilon)}_{t,0}(\boldsymbol{x}) - \widetilde{\boldsymbol{\theta}}_{t,0}(\boldsymbol{x})\big)_i\big|, \quad i = 1, 2.$$

For $i = 1$, noting that

$$|F^{(\varepsilon)}_1(t, \boldsymbol{x}) - F_1(t, \boldsymbol{x})| \leqslant \kappa_1(1 + \varepsilon),$$

by definition we have

$$\begin{aligned}
\ell_1(t) &\leqslant \int_0^t |F^{(\varepsilon)}_1(r, \boldsymbol{\theta}^{(\varepsilon)}_r(\boldsymbol{x})) - F^{(1)}_1(r, \widetilde{\boldsymbol{\theta}}_r(\boldsymbol{x}))|\mathrm{d}r \\
&\leqslant \int_0^t |F^{(\varepsilon)}_1(r, \boldsymbol{\theta}^{(\varepsilon)}_r(\boldsymbol{x})) - F^{(1)}_1(r, \boldsymbol{\theta}^{(\varepsilon)}_r(\boldsymbol{x}))|\mathrm{d}r \\
&\quad + \int_0^t |F^{(1)}_1(r, \boldsymbol{\theta}^{(\varepsilon)}_r(\boldsymbol{x})) - F^{(1)}_1(r, \widetilde{\boldsymbol{\theta}}_r(\boldsymbol{x}))|\mathrm{d}r \\
&\leqslant 2\kappa_1 t + \|\nabla F^{(1)}_1\|_\infty \int_0^t (\ell_1(r) + \ell_2(r))\mathrm{d}r,
\end{aligned}$$

which implies by Gronwall's inequality

$$\ell_1(t) \lesssim t + \int_0^t \ell_2(r)\mathrm{d}r. \tag{4.4}$$

For $i = 2$, note that

$$|F^{(\varepsilon)}_2(t, \boldsymbol{x}) - F^{(\varepsilon)}_2(t, \boldsymbol{y})| \leqslant \kappa_2(|x_1 - y_1| + |x_2 - y_2|^{(1+\gamma)/3} + |x_2 - y_2|)$$

and

$$|F_2^{(\varepsilon)}(t,\boldsymbol{x}) - F_2(t,\boldsymbol{x})| \leqslant \kappa_2(\varepsilon^{(1+\gamma)/3} + 2\varepsilon).$$

Below we fix $t \in (0,T]$ and $\varepsilon \leqslant t^{3/2}$. By definition we have for all $s \in [0,t]$,

$$\begin{aligned}
\ell_2(s) &\leqslant \int_0^s |F_2^{(\varepsilon)}(r,\boldsymbol{\theta}_r^{(\varepsilon)}(\boldsymbol{x})) - [F_2(r,\cdot) * \rho_{r^{3/2}}](\widetilde{\boldsymbol{\theta}}_r(\boldsymbol{x}))|\mathrm{d}r \\
&\leqslant \int_0^s |F_2^{(\varepsilon)}(r,\boldsymbol{\theta}_r^{(\varepsilon)}(\boldsymbol{x})) - F_2^{(\varepsilon)}(r,\widetilde{\boldsymbol{\theta}}_r(\boldsymbol{x}))|\mathrm{d}r \\
&\quad + \int_0^s |F_2^{(\varepsilon)}(r,\widetilde{\boldsymbol{\theta}}_r(\boldsymbol{x})) - [F_2(r,\cdot) * \rho_{r^{3/2}}](\widetilde{\boldsymbol{\theta}}_r(\boldsymbol{x}))|\mathrm{d}r \\
&\lesssim \int_0^s \big(\ell_1(r) + \ell_2(r)^{\frac{1+\gamma}{3}} + \ell_2(r)\big)\mathrm{d}r + st^{(1+\gamma)/2} + \int_0^s r^{(1+\gamma)/2}\mathrm{d}r \\
&\lesssim t^{(3+\gamma)/2} + \int_0^s \big(\ell_2(r)^{\frac{1+\gamma}{3}} + \ell_2(r)\big)\mathrm{d}r.
\end{aligned}$$

By Lemma 2.1, we obtain

$$\sup_{s\in[0,t]} \ell_2(s) \lesssim t^{3/2},$$

which together with (4.4) yields (4.3). $\square$

Now by (4.2) and (4.3), there is a constant $C_0 > 0$ such that for any nonnegative $f \in C_b(\mathbb{R}^{2d})$ and $\varepsilon \leqslant (t-s)^{3/2}$,

$$C_0^{-1} \int_{\mathbb{R}^{2d}} g_{\lambda_0^{-1}}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big) f(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \leqslant \mathbb{E} f(\mathbf{X}^{\varepsilon}_{t,s}(\boldsymbol{x})) \leqslant C_0 \int_{\mathbb{R}^{2d}} g_{\lambda_0}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big) f(\boldsymbol{y})\mathrm{d}\boldsymbol{y},$$

which together with (4.1) yields that

$$C_0^{-1} \int_{\mathbb{R}^{d}} g_{\lambda_0^{-1}}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big) f(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \leqslant \mathbb{E} f(\mathbf{X}_{t,s}(\boldsymbol{x})) \leqslant C_0 \int_{\mathbb{R}^{d}} g_{\lambda_0}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big) f(\boldsymbol{y})\mathrm{d}\boldsymbol{y}.$$

In particular, this implies that $\mathbf{X}_{t,s}(\boldsymbol{x})$ has a density $p(s,\boldsymbol{x};t,\boldsymbol{y})$ having lower and upper bound as in (1.16). This proves point *(i)* of Theorem 1.1.

Similarly, we derive from Propositions 3.10, 3.11 and 3.12 that under $(\mathbf{H}^{\gamma}_{\mathbf{F}})$ and $(\mathbf{H}^{\gamma}_{\sigma})$,

$$\sup_{\varepsilon\leqslant(t-s)^{3/2}} |\nabla_{x_1} p_\varepsilon(s,\boldsymbol{x};t,\boldsymbol{y})| \leqslant C_1(t-s)^{-\frac{1}{2}} g_{\lambda_1}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big), \tag{4.5}$$

under $(\mathbf{H}^{\gamma}_{\mathbf{F}})$, $(\mathbf{H}^{\gamma}_{\sigma})$ with (1.19),

$$\sup_{\varepsilon\leqslant(t-s)^{3/2}} \big|\nabla^2_{x_1} p_\varepsilon(s,\boldsymbol{x};t,\boldsymbol{y})\big| \leqslant C_2(t-s)^{-1} g_{\lambda_1}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big), \tag{4.6}$$

and under $(\mathbf{H}^{\gamma}_{\mathbf{F}})$, $(\mathbf{H}^{\gamma}_{\sigma})$ with (1.22), (1.23),

$$\sup_{\varepsilon\leqslant(t-s)^{3/2}} |\nabla_{x_2} p_\varepsilon(s,\boldsymbol{x};t,\boldsymbol{y})| \leqslant C_3(t-s)^{-\frac{3}{2}} g_{\lambda_3}\big(t-s, \widetilde{\boldsymbol{\theta}}_{t,s}(\boldsymbol{x}) - \boldsymbol{y}\big), \tag{4.7}$$

where in the above equations (4.5)-(4.7), the constants C_1-C_3 only depend on Θ_T. Uniform Hölder controls in $\boldsymbol{x}$ also follow from the previously recalled Propositions. Equicontinuity w.r.t. the variable $\boldsymbol{y}$ could be established similarly.

Then, from the Ascoli-Arzelà theorem, one can find a subsequence ε_k such that for each $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2d}$,

$$\nabla_{x_1}^j p_{\varepsilon_k}(s, \boldsymbol{x}; t, \boldsymbol{y}) \to \nabla_{x_1}^j p(s, \boldsymbol{x}; t, \boldsymbol{y}), \quad j = 0, 1, 2, \ \nabla_{x_2} p_{\varepsilon_k}(s, \boldsymbol{x}; t, \boldsymbol{y}) \to \nabla_{x_2} p(s, \boldsymbol{x}; t, \boldsymbol{y}).$$

The gradient estimates follow, under the previously recalled additional assumptions when needed, from (4.5)-(4.7).

Acknowledgements

For the second author this research has been partially funded by RSF grant 20-11-20119 for this work which began in May 2021.

References

[1] Barucci, E., Polidoro, S., and Vespri, V. Some results on partial differential equations and asian options. *Math. Models Methods Appl. Sci 3* (2001), 475–497.

[2] Chaudru de Raynal, P., Honoré, I., and Menozzi, S. Strong Uniqueness for a chain of degenerate Kolmogorov equations. *arXiv:1810.12225* (2018).

[3] Chaudru de Raynal, P.-E. Strong existence and uniqueness for degenerate SDE with Höder drift. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 53*, 1 (Feb. 2017), 259–286.

[4] Chaudru de Raynal, P.-E. Weak regularization by stochastic drift: result and counter example. *Discrete and Continuous Dynamical Systems (Series A) 38–3* (2018), 1269–1291.

[5] Chaudru de Raynal, P.-E., Honoré, I., and Menozzi, S. Sharp Schauder estimates for some Degenerate Kolmogorov Equations. *Annali Scuola Normale Superiore di Pisa 22-3* (2021), 989–1089.

[6] Chaudru de Raynal, P.-E., and Menozzi, S. Regularization effects of a noise propagating through a chain of differential equations: An almost sharp result. *Transactions of the American Mathematical Society 375–01* (2022), 1–45.

[7] Chen, Z.-Q., and Zhang, X. Propagation of regularity in L^p-spaces for Kolmogorov-type hypoelliptic operators. *J. Evol. Equ. 19*, 4 (2019), 1041–1069.

[8] Cibelli, G., Polidoro, S., and Rossi, F. Sharp estimates for Geman-Yor processes and applications to arithmetic average Asian options. *J. Math. Pures Appl. (9) 129* (2019), 87–130.

[9] Delarue, F., and Menozzi, S. Density estimates for a random noise propagating through a chain of differential equations. *Journal of Functional Analysis 259–6* (2010), 1577–1630.

[10] Fedrizzi, E., Flandoli, F., Priola, E., and Vovelle, J. Regularity of stochastic kinetic equations. *Electronic Journal of Probability 22* (2017), ♯ 48.

[11] Flandoli, F. *Random perturbation of PDEs and fluid dynamic models*, vol. 2015 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2011. Lectures from the 40th Probability Summer School held in Saint-Flour, 2010.

[12] Friedman, A. *Partial differential equations of parabolic type.* Prentice-Hall, 1964.

[13] Gadat, S., and Panloup, F. Long time behaviour and stationary regime of memory gradient diffusions. *Ann. Inst. Henri Poincaré Probab. Stat. 50*, 2 (2014), 564–601.

[14] Hao, Z., Wu, M., and Zhang, X. Schauder estimates for nonlocal kinetic equations and applications. *J. Math. Pures Appl. (9) 140* (2020), 139–184.

[15] Hérau, F., and Nier, F. Isotropic hypoellipticity and trend to equilibrium for the Fokker-planck equation with a high-degree potential. *Arch. Ration. Mech. Anal. 171–2* (2004), 151–218.

[16] Hörmander, L. Hypoelliptic second order differential operators. *Acta. Math. 119* (1967), 147–171.

[17] Huang, L., Menozzi, S., and Priola, E. L^p Estimates For Degenerate Non-Local Kolmogorov Operators. *Journal de Math. Pures et Appl. 121* (2019), 162–215.

[18] Imbert, C., and Mouhot, C. The Schauder estimate in kinetic theory with application to a toy nonlinear model. *Ann. H. Lebesgue 4* (2021), 369–405.

[19] Kolmogorov, A. N. Zufällige Bewegungen (zur Theorie der Brownschen Bewegung). *Ann. of Math. 2-35* (1934), 116–117.

[20] Konakov, V., Menozzi, S., and Molchanov, S. Explicit parametrix and local limit theorems for some degenerate diffusion processes. *Annales de l'Institut Henri Poincaré, Série B 46–4* (2010), 908–923.

[21] Kusuoka, S., and Stroock, D. Applications of the Malliavin calculus. III. *J. Fac. Sci. Univ. Tokyo Sect. IA Math 34* (1987), 391–442.

[22] Lanconelli, A., and Pascucci, A. Nash estimates and upper bounds for non-homogeneous Kolmogorov equations. *Potential Anal. 47*, 4 (2017), 461–483.

[23] Lanconelli, A., Pascucci, A., and Polidoro, S. Gaussian lower bounds for non-homogeneous Kolmogorov equations with measurable coefficients. *J. Evol. Equ. 20*, 4 (2020), 1399–1417.

[24] Lanconelli, E., and Polidoro, S. On a class of hypoelliptic evolution operators. *Rend. Sem. Mat. Univ. Politec. Torino 52–1* (1994), 29–63.

[25] Lunardi, A. Schauder estimates for a class of degenerate elliptic and parabolic operators with unbounded coefficients in $\mathbf{R}^n$. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) 24*, 1 (1997), 133–164.

[26] Marino, L. Schauder estimates for degenerate stable Kolmogorov equations. *Bull. Sci. Math. 162* (2020), 102885, 98.

[27] Mattingly, J., Stuart, J., and Higham, D. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stoch. Proc. Appl. 101–2* (2002), 185–232.

[28] McKean, H. P., and Singer, I. M. Curvature and the eigenvalues of the Laplacian. *J. Differential Geometry 1* (1967), 43–69.

[29] Menozzi, S. Parametrix techniques and martingale problems for some degenerate Kolmogorov equations. *Electronic Communications in Probability 17* (2011), 234–250.

[30] Menozzi, S. Martingale problems for some degenerate Kolmogorov equations. *Stoc. Proc. Appl. 128-3* (2018), 756–802.

[31] Menozzi, S., Pesce, A., and Zhang, X. Density and gradient estimates for non degenerate brownian sdes with unbounded measurable drift. *Journal of Differential Equations 272* (2021), 330 – 369.

[32] Menozzi, S., and Zhang, X. Heat kernel of supercritical sdes with unbounded drifts. *arXiv:2012.14775. To Appear in Journal de l'Ecole Polytechnique* (2020).

[33] Pascucci, A., and Pesce, A. On stochastic Langevin and Fokker-Planck equations: The two-dimensional case. *J. Differential Equations 310* (2022), 443–483.

[34] Pascucci, A., and Polidoro, S. Harnack inequalities and Gaussian estimates for a class of hypoelliptic operators. *Trans. Amer. Math. Soc. 358–11* (2006), 4873–4893.

[35] Pigato, P. Density estimates and short-time asymptotics for a hypoelliptic diffusion process. *Stochastic Processes and Their Applications 145* (2022), 117–142.

[36] Soize, C. *The Fokker-Planck equation for stochastic dynamical systems and its explicit steady state solutions.* Series on Advances in Mathematics for Applied Sciences, 17. World Scientific Publishing Co., Inc., River Edge, NJ, 1994.

[37] Sonin, I. M. A class of degenerate diffusion processes (russian). *Teor. Verojatnost. i Primenen 12* (1967), 540–547.

[38] Stroock, D., and Varadhan, S. *Multidimensional diffusion processes.* Springer-Verlag Berlin Heidelberg New-York, 1979.

[39] Talay, D. Stochastic Hamiltonian dissipative systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Processes and Related Fields 8–2* (2002), 163–198.

[40] Wang, F., and Zhang, X. Degenerate SDE with Holder-Dini drift and non-Lipschitz noise coefficient. *SIAM J. Math. Anal. 48*, 3 (2016), 2189–2226.

[41] Weber, M. The fundamental solution of a degenerate partial differential equation of parabolic type. *Trans. Amer. Math. Soc. 71* (1951), 24–37.

[42] Zhang, X. A variational representation for random functionals on abstract Wiener spaces. *J. Math. Kyoto Univ. 49*, 3 (2009), 475–490.

[43] Zhang, X. Stochastic Hamiltonian flows with singular coefficients. *Sci. China Math. 61*, 8 (2018), 1353–1384.

[44] Zhang, X. Second order McKean-Vlasov SDEs and kinetic Fokker-Planck-Kolmogorov equations. *arXiv:2109.01273* (2021).