Supplementary Material

This supplementary material contains the proofs and results omitted from the main body. In Appendix A we recall the appropriate version of the Stokes’ theorem and discuss its applicability for Lipschitz functions on $B_2^d$. In Appendix B we provide the proof of Lemma 3. Finally, in Appendix C we provide the proofs of Theorems 1, 2, 3, 4.

Additional notation  For two functions $g, \eta : \mathbb{R}^d \to \mathbb{R}$, we denote by $\eta * g$ their convolution defined point-wise for $x \in \mathbb{R}^d$ as

$$(\eta * g)(x) = \int_{\mathbb{R}^d} \eta(x - x')g(x') \, dx'.$$

The standard mollifier $\eta_\varepsilon : \mathbb{R}^d \to \mathbb{R}$ is defined as $\eta_\varepsilon(x) = \varepsilon^{-d} \eta_1(x/\varepsilon)$ for $\varepsilon > 0$ and $x \in \mathbb{R}$, where $\eta_1 : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\eta_1(x) = \begin{cases} C \exp \left( \frac{1}{|x|^{1 - \tau}} \right) & \text{if } |x|_2 \leq 1, \\ 0 & \text{otherwise} \end{cases},$$

with $C$ chosen so that $\int_{\mathbb{R}^d} \eta_1(x) \, dx = 1$.

A Integration by parts

We first recall the following result that can be found in [34, Section 13.3.5, Exercise 14a].

**Theorem 5** (Integration by parts in a multiple integral). Let $D$ be an open connected subset of $\mathbb{R}^d$ with a piecewise smooth boundary $\partial D$ oriented by the outward unit normal $n = (n_1, \ldots, n_d)^	op$. Let $g$ be a continuously differentiable function in $D \cup \partial D$. Then

$$\int_D \nabla g(u) \, du = \int_{\partial D} g(\zeta)n(\zeta) \, dS(\zeta).$$

**Remark 2.** We refer to [34, Section 12.3.2, Definitions 4 and 5] for the definition of piecewise smooth surfaces and their orientations respectively.

The idea of using the instance of Theorem 5 (also called Stokes’ theorem) with $D = B_2^d$ to obtain $\ell_2$-randomized estimators of the gradient belongs to Nemirovsky and Yudin [22]. It was further used in several papers [5, 16, 31, 33] to mention just a few. Those papers were referring to [22] but [22] did not provide an exact statement of the result (nor a reference) and only tossed the idea in a discussion. However, the classical analysis formulation as presented in Theorem 5 does not apply to Lipschitz continuous functions that were considered in [5, 16, 31, 33]. We are not aware of whether its extension to Lipschitz continuous functions, though rather standard, is proved in the literature.

In this paper, we apply Theorem 5 with the $\ell_1$-ball $D = B_1^d$. Our aim in this section is to provide a variant of Theorem 5 applicable to a Lipschitz continuous function $g : \mathbb{R}^d \to \mathbb{R}$, which is not necessarily continuously differentiable on $D \cup \partial D = B_1^d \cup \partial B_1^d$. To this end, we will go through the argument of approximating $g$ by $C^\infty(\Omega)$ functions, where $\Omega \subset \mathbb{R}^d$ is an open bounded connected subset of $\mathbb{R}^d$ such that $D \cup \partial D \subset \Omega$. Let $g_n = \eta_{1/n} * g$, where $\eta_{1/n}$ is the standard mollifier. Let $g : \mathbb{R}^d \to \mathbb{R}$ be a function satisfying the Lipschitz condition w.r.t. the $\ell_1$-norm: $|g(u) - g(u')| \leq L||u - u'||_1$. Since $g$ is continuous in $\Omega$ and, by construction $D \cup \partial D \subset \Omega$, then using basic properties of mollification [see e.g., 15, Theorem 4.1 (ii)] we have $g_n \to g$ uniformly on $D \cup \partial D$ (in particular, uniformly on $\partial D$). Furthermore, let $\nabla g$ be the gradient of $g$, which by Rademacher theorem [see e.g., 15, Theorem 3.2] is well defined almost everywhere w.r.t. the Lebesgue measure and

$$\|\nabla g(u)\|_\infty \leq L \quad \text{a.e.}$$
It follows that \( \frac{\partial g}{\partial u_j} \) is absolutely integrable on \( \Omega \) for any \( j \in [d] \). Furthermore, since
\[
\frac{\partial g_n}{\partial u_j} = \eta_{1/n} \ast \left( \frac{\partial g}{\partial u_j} \right),
\]
we can apply [15, Theorem 4.1 (iii)] that yields
\[
\int_D \| \nabla g_n(u) - \nabla g(u) \|_2 \, du \to 0.
\]
Combining the above remarks we obtain that the result of Theorem 5 is valid for functions \( g \) that are Lipschitz continuous w.r.t. the \( \ell_1 \)-norm. Thus, it is also valid when the Lipschitz condition is imposed w.r.t. any \( \ell_q \)-norm with \( q \in [1, \infty] \). Specifying this conclusion for the particular case \( D = B_1^d \), we obtain the following theorem.

**Theorem 6.** Let the function \( g : \mathbb{R}^d \to \mathbb{R} \) be Lipschitz continuous w.r.t. the \( \ell_q \)-norm with \( q \in [1, \infty] \). Then
\[
\int_{B_1^d} \nabla g(u) \, du = \frac{1}{\sqrt{d}} \int_{\partial B_1^d} g(\zeta) \, \text{sign}(\zeta) \, dS(\zeta),
\]
where \( \nabla g(\cdot) \) is defined up to a set of zero Lebesgue measure by the Rademacher theorem.

## B Proof of Lemma 3

To prove Lemma 3, we first recall the weighted Poincaré inequality for the univariate exponential measure (mean 0 and scale parameter 1 Laplace distribution).

**Lemma 5** (Lemma 2.1 in [9]). Let \( W \) be mean 0 and scale parameter 1 Laplace random variable. Let \( g : \mathbb{R} \to \mathbb{R} \) be continuous almost everywhere differentiable function such that
\[
\mathbb{E}[|g(W)|] < \infty \quad \text{and} \quad \mathbb{E}[|g'(W)|] < \infty \quad \text{and} \quad \lim_{|w| \to \infty} g(w) \exp(-|w|) = 0,
\]
then,
\[
\mathbb{E}[(g(W) - \mathbb{E}[g(W)])^2] \leq 4\mathbb{E}[(g'(W))^2].
\]

We are now in a position to prove Lemma 3. The proof is inspired by [7, Lemma 2].

**Proof of Lemma 3.** Throughout the proof, we assume without loss of generality that \( \mathbb{E}[G(\zeta)] = 0 \). Indeed, if it is not the case, we use the result for the centered function \( \tilde{G}(\zeta) = G(\zeta) - \mathbb{E}[G(\zeta)] \), which has the same gradient.

First, consider the case of continuously differentiable \( G \). Let \( W = (W_1, \ldots, W_d) \) be a vector of i.i.d. mean 0 and scale parameter 1 Laplace random variables and define \( T(w) = w/\|w\|_1 \). Introduce the notation
\[
F(w) \triangleq \|w\|_1^{1/2}G(T(w)).
\]
Lemma 1 in [30] asserts that, for \( \zeta \) uniformly distributed on \( \partial B_1^d \),
\[
T(W) \overset{d}{=} \zeta \quad \text{and} \quad T(W) \text{ is independent of } \|W\|_1.
\]
In particular,
\[
\text{Var}(F(W)) = d \text{Var}(G(\zeta)).
\]
Using the Efron-Stein inequality [see e.g., 11, Theorem 3.1] we obtain
\[
\text{Var}(F(W)) \leq \sum_{i=1}^d \mathbb{E}[\text{Var}_i(F)],
\]
where
\[
\text{Var}_i(F) = \mathbb{E}\left[ (F(W) - \mathbb{E}[F(W) | W^{-i}])^2 \mid W^{-i} \right]
\]
with $W^{-i} \triangleq \{W_1, \ldots, W_{i-1}, W_{i+1}, \ldots, W_d\}$. Note that on the event $\{W^{-i} \neq 0\}$ (whose complement has zero measure), the function

$$w \mapsto F(W_1, \ldots, W_{i-1}, w, W_{i+1}, \ldots, W_d)$$

satisfies the assumptions of Lemma 5. Thus,

$$d \text{Var}(G(\zeta)) = \text{Var}(F(W)) \leq 4 \sum_{j=1}^{d} E \left[ \left( \frac{\partial F}{\partial w_j}(W) \right)^2 \right] = 4E\|\nabla F(W)\|_2^2. \quad (5)$$

In order to compute $\nabla F(W)$, we observe that for every $i \neq j \in [d]$ we have for all $w \neq 0$ such that $w_i, w_j \neq 0$

$$\frac{\partial T_i}{\partial w_j}(w) = -\frac{w_i \text{sign}(w_j)}{\|w\|_1^2} \quad \text{and} \quad \frac{\partial T_i}{\partial w_i}(w) = \frac{1}{\|w\|_1} - \frac{w_i \text{sign}(w_j)}{\|w\|_1^2}.$$

Thus, the Jacobi matrix of $T(w)$ has the form

$$J_T(w) = \frac{1}{\|w\|_1} - \frac{w(\text{sign}(w))^T}{\|w\|_1^2} = \frac{1}{\|w\|_1} \left( I - T(w)(\text{sign}(w))^T \right) .$$

It follows that almost surely

$$\nabla F(W) = \frac{1}{2\|W\|_1^{1/2}}G(T(W))\text{sign}(W) + \frac{1}{\|W\|_1^{1/2}} \left( I - T(W)(\text{sign}(W))^T \right) \nabla G(T(W)) .$$

Observe that since $(\text{sign}(W), T(W)) = 1$ almost surely, we have

$$(\text{sign}(W))^T \left( I - T(W)(\text{sign}(W))^T \right) \nabla G(T(W)) = 0 \quad \text{almost surely.}$$

The above two equations imply that almost surely

$$4\|\nabla F(W)\|_2^2 = \frac{d}{\|W\|_1}G^2(T(W)) + \frac{4}{\|W\|_1} \left\| \left( I - T(W)(\text{sign}(W))^T \right) \nabla G(T(W)) \right\|_2^2 \leq \frac{d}{\|W\|_1}G^2(T(W)) + \frac{4}{\|W\|_1} \|\nabla G(T(W))\|_2^2 \left( 1 + \sqrt{d\|T(W)\|_2^2} \right) ,$$

where we used the fact that the operator norm of $I - ab^T$ is not greater than $1 + \|a\|_2\|b\|_2$. Combining the above bound with (5), and using the facts that $E[\|W\|_1^{-1}] = \frac{d}{d-1} - 1$, $E[G(T(W))] = E[G(\zeta)] = 0$ and the independence of $\|W\|_1$ and $T(W)$ (cf. (4)) yields

$$d \left( 1 - \frac{1}{d-1} \right) \text{Var}(G(\zeta)) \leq \frac{4}{d-1} E \left[ \|\nabla G(T(W))\|_2^2 \left( 1 + \sqrt{d\|T(W)\|_2^2} \right) \right] .$$

Rearranging, we deduce the first claim of the lemma since $T(W) \overset{d}{=} \zeta$.

To prove the second statement of the lemma regarding Lipschitz functions, it is sufficient to apply the first one to $G_n$—the sequence of smoothed versions of $G$ such that $G_n \in C^\infty(\mathbb{R})$ and

$$G_n \to G ,$$

uniformly on every compact subset, and $\sup_{n \geq 1} \|\nabla G_n(x)\|_2 \leq L$ for almost all $x \in \mathbb{R}^d$. A sequence $G_n$ satisfying these properties can be constructed by standard mollification due to the fact that $G$ is Lipschitz continuous [see e.g., 15, Theorem 4.2]. Finally, to obtain the value $E\|T(W)\|_2^2 = E\|\zeta\|_2^2$ we use Lemma 6 below.

**Lemma 6.** Let $\zeta$ be distributed uniformly on $\partial B_1^d$. Then, $E\|\zeta\|_2^2 = \frac{2}{d+1}$.

**Proof.** We use the same tools as in the proof of Lemma 2. Let $W = (W_1, \ldots, W_d)$ be a vector of i.i.d. random variables following the Laplace distribution with mean 0 and scale parameter 1. By (4) we have that $\zeta \overset{d}{=} \frac{W}{\|W\|_1}$ and $\zeta$ is independent of $\|W\|_1$. Therefore,

$$E\|\zeta\|_2^2 = \frac{E\|W\|_1^2}{E\|W\|_1^2} . \quad (6)$$
Here,

$$E \| W \|_2^2 = \sum_{j=1}^{d} E[W_j^2] = dE[W_1^2] = 2d.$$  \hfill (7)

Furthermore, $\| W \|_1$ follows the Erlang distribution with parameters $(d, 1)$, which implies

$$E \| W \|_1^2 = \frac{1}{\Gamma(d)} \int_0^\infty x^{d+1} \exp(-x) \, dx = \frac{\Gamma(d+2)}{\Gamma(d)}.$$  \hfill (8)

The lemma follows by combining (6) – (8).

\[\square\]

\section{C Upper bounds}

The proofs of Theorems 1, 2, 3, 4 resemble each other. They only differ in the ways of handling the variance terms depending on $\| g_t \|_p^2$, and in the choice of parameters. For this reason, we suggest the interested reader to follow the proofs in a linear manner starting from the next paragraph.

\subsection{Common part of the proofs of Theorems 1, 2}

We start with the part of the proofs that is common for Theorems 1, 2. Fix some $x \in \Theta$. Due to Assumption 1, we can use Lemma 1, which implies

$$E \left[ \sum_{t=1}^{T} (E[g_t | x_t], x_t - x) \right] = E \left[ \sum_{t=1}^{T} \langle \nabla f_{t,h}(x_t), x_t - x \rangle \right] \geq E \left[ \sum_{t=1}^{T} (f_{t,h}(x_t) - f_{t,h}(x)) \right],$$

where $f_{t,h}(x) = E[f_t(x + hU)]$ with $U$ uniformly distributed on $B_1^d$. Furthermore, by the approximation property derived in Lemma 1 and the standard bound on the cumulative regret of dual averaging algorithm [see e.g., 26, Corollary 7.9] we deduce that

$$E \left[ \sum_{t=1}^{T} (f_t(x_t) - f_t(x) \right] \leq E \left[ \sum_{t=1}^{T} (E[g_t | x_t], x_t - x) \right] + Lb_t(d) \sum_{t=1}^{T} h_t \leq \frac{R^2}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} E \| g_t \|_p^2 + Lb_t(d) \sum_{t=1}^{T} h_t,$$  \hfill (9)

where in the last inequality we used the identity $\eta_1 = \ldots = \eta_T = \eta$. The results of Theorems 1, 2 follow from the bound (9) as detailed below.

\subsection{Proof of Theorem 1}

Here $h_1 = \ldots = h_T = h$, and we work under Assumption 2. In this case, bounding $E \| g_t \|_p^2$ in (9) via Lemma 4 yields

$$E \left[ \sum_{t=1}^{T} (f_t(x_t) - f_t(x) \right] \leq \frac{R^2}{\eta} + 6(1 + \sqrt{2})^2 L^2 \cdot \eta Td^{1 + \frac{1}{2 \gamma} - \frac{h}{2}} + LhTb_t(d).$$

Minimizing the the right hand side of the above inequality over $\eta > 0$ and substituting $\eta = \frac{R}{L(\sqrt{6} + \sqrt{12})} \sqrt{d^{1 - \frac{2}{2 \gamma} + \frac{h}{2}}}$ we deduce that

$$E \left[ \sum_{t=1}^{T} (f_t(x_t) - f_t(x) \right] \leq 2 \left( \sqrt{6} + \sqrt{12} \right) RLd^{1 + \frac{1}{2 \gamma} - \frac{h}{2}} + LhTb_t(d).$$

Taking $h \leq \frac{7R}{100b_t(d)} d^{1 + \frac{1}{2 \gamma} - \frac{h}{2}}$ makes negligible the second summand in the above bound. This concludes the proof.

\subsection{Proof of Theorem 2}

Here again $h_1 = \ldots = h_T = h$, but we work under Assumption 3. Then, bounding $E \| g_t \|_p^2$ in (9) via Lemma 4 yields

$$E \left[ \sum_{t=1}^{T} (f_t(x_t) - f_t(x) \right] \leq \frac{R^2}{\eta} + \eta T \left( \frac{d^{1 - \frac{2}{2 \gamma}} \sigma^2}{h^2} + 6 \left( 1 + \sqrt{2} \right)^2 L^2 d^{1 + \frac{1}{2 \gamma} - \frac{h}{2}} \right) + LhTb_t(d).$$
Minimizing the right hand side of the above inequality over $\eta > 0$ and substituting the optimal value

$$\eta = \frac{R}{\sqrt{T}} \left( \frac{d^{1-\frac{2}{p}} \sigma^2}{2h^2} + 6 \left(1 + \sqrt{2} \right)^2 L^2 d^{1+\frac{2}{\alpha^2} - \frac{2}{h}} \right)^{-\frac{1}{2}},$$

results in the following upper bound on the regret

$$\mathbb{E} \left[ \sum_{t=1}^T (f_t(x_t) - f_t(x)) \right] \leq 2R\sqrt{T} \left( \frac{d^{1-\frac{2}{p}} \sigma^2}{2h^2} + 6 \left(1 + \sqrt{2} \right)^2 L^2 d^{1+\frac{2}{\alpha^2} - \frac{2}{h}} \right)^{\frac{1}{2}} + LhTb_\eta(d)$$

$$\leq 2 \left( \sqrt{6} + \sqrt{12} \right) RL \sqrt{T d^{1+\frac{2}{\alpha^2} - \frac{2}{h}}} + 2.4\sqrt{R\sigma T^2} + LhTb_\eta(d),$$

where for the last inequality we used the fact that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b \geq 0$. Minimizing over $h > 0$ the last expression and substituting the optimal value $h = \left( \frac{2RL}{Lb_\eta(d)} \right)^{\frac{1}{2}} T^{-\frac{1}{2}} d^{-\frac{1}{2h}}$ we get

$$\mathbb{E} \left[ \sum_{t=1}^T (f_t(x_t) - f_t(x)) \right] \leq 11.9RL \sqrt{T d^{1+\frac{2}{\alpha^2} - \frac{2}{h}}} + 2.4\sqrt{R\sigma T^2} + Lb_\eta(d) d^{\frac{1}{2h}}. \quad \Box$$

**Common part of the proofs of Theorems 3, 4.** Here, we state the common parts of the proofs for Theorems 3, 4. Similar to the first inequality in (9), we have

$$\mathbb{E} \left[ \sum_{t=1}^T (f_t(x_t) - f_t(x)) \right] \leq \mathbb{E} \left[ \sum_{t=1}^T (g_t, x_t - x) \right] + Lb_\eta(d) \sum_{t=1}^T h_t.$$  

Note that without loss of generality, we can assume that $\sum_{k=1}^t \| g_k \|_p^2 \neq 0$, for all $t \geq 1$. This is a consequence of the fact that if $\sum_{k=1}^t \| g_k \|_p^2 = 0$, then the first term on the r.h.s. of the above inequality will be zero up to round $t$. Thus, we can erase these iterates from the cumulative regret, only paying the bias term for those rounds. In what follows we essentially use [27, Corollary 1], which we re-derive for the sake of clarity. Assume that $\eta_t = \frac{\lambda}{\sqrt{\sum_{k=1}^t \| g_k \|_p^2}}$ for $t \in \{2, \ldots, T\}$ and $\lambda > 0$. Then, applying [27, Theorem 1] we deduce that

$$\mathbb{E} \left[ \sum_{t=1}^T (f_t(x_t) - f_t(x)) \right] \leq \left( \frac{R^2}{\lambda} + 2.75 \lambda \right) \mathbb{E} \left[ \sum_{t=1}^T \| g_t \|_p^2 \right]$$

$$+ 3.5D \cdot \mathbb{E} [\max_{t \in [T]} \| g_t \|_p^2] + Lb_\eta(d) \sum_{t=1}^T h_t,$$

where we introduced $D = \sup_{u, w \in \Theta} \| u - w \|_p$. By [27, Proposition 1], we have $D \leq \sqrt{8} R$. Moreover, by Jensen’s inequality, using the rough bound $\mathbb{E} [\max_{t \in [T]} \| g_t \|_p^2] \leq \sqrt{\sum_{t=1}^T \mathbb{E} \left[ \| g_t \|_p^2 \right]},$ and substituting $\lambda = \frac{R}{\sqrt{2.75}}$, we deduce that

$$\mathbb{E} \left[ \sum_{t=1}^T (f_t(x_t) - f_t(x)) \right] \leq \left( 2\sqrt{2.75} + 3.5\sqrt{8} \right) R \sqrt{\sum_{t=1}^T \mathbb{E} \left[ \| g_t \|_p^2 \right]} + Lb_\eta(d) \sum_{t=1}^T h_t.$$  

(10)

Proofs of Theorems 3, 4 provided below follow from the above inequality by properly selecting $h_t > 0$.

**Proof of Theorem 3.** The bound of Lemma 4 under Assumption 2 applied to (10) yields

$$\mathbb{E} \left[ \sum_{t=1}^T (f_t(x_t) - f_t(x)) \right] \leq 2 \left( 2\sqrt{2.75} + 3.5\sqrt{8} \right) \left( \sqrt{3} + \sqrt{6} \right) RL \sqrt{T d^{1+\frac{2}{\alpha^2} - \frac{2}{h}}} + Lb_\eta(d) \sum_{t=1}^T h_t$$

$$\leq 110.53 \cdot RL \sqrt{T d^{1+\frac{2}{\alpha^2} - \frac{2}{h}}} + Lb_\eta(d) \sum_{t=1}^T h_t.$$
Taking \( h_t \leq \frac{7R}{200b_d(d)\sqrt{t}} d^{1+\frac{3}{\sqrt{d}}} - \frac{1}{p} \) makes negligible the last summand in the above bound. This concludes the proof.

**Proof of Theorem 4.** Using (10), the bound of Lemma 4 under Assumption 3 and the fact that \( \sqrt{a} + b \leq \sqrt{a} + \sqrt{b} \) for \( a, b \geq 0 \), we deduce that

\[
E \left[ \sum_{t=1}^{T} (f_t(x_t) - f_t(x)) \right] \leq \left( 2\sqrt{2.75} + 3.5\sqrt{8} \right) \frac{R}{h_t^2} \left( \sum_{t=1}^{T} \frac{d^{1-\frac{3}{\sqrt{d}}} \sigma^2}{h_t^2} + 12(1 + \sqrt{2})^2 L^2 T \cdot d^{1+\frac{3}{\sqrt{d}}} - \frac{1}{p} \right)^{\frac{1}{2}}
\]

\[
+ Lb_q(d) \left( \sum_{t=1}^{T} h_t \right)
\]

\[
\leq 110.6 \cdot RL \sqrt{T \cdot d^{1+\frac{3}{\sqrt{d}}} - \frac{1}{p}} + 13.3R \cdot d^{2-\frac{1}{p}} \sigma \left( \sum_{t=1}^{T} \frac{1}{h_t^2} \right)^{\frac{1}{2}}
\]

\[
+ Lb_q(d) \left( \sum_{t=1}^{T} h_t \right)
\]

Since \( h_t = \left( 6.65\sqrt{6} \cdot \frac{R}{h_t(d)} \right)^{\frac{1}{2}} t^{-\frac{1}{4}} d^{1-\frac{1}{\sqrt{d}}} \) and \( \sum_{t=1}^{T} t^{\frac{1}{2}} \leq \frac{2}{3} T^{\frac{3}{2}} \) and \( \sum_{t=1}^{T} t^{-\frac{1}{4}} \leq \frac{4}{3} T^{\frac{1}{4}} \), we get

\[
E \left[ \sum_{t=1}^{T} (f_t(x_t) - f_t(x)) \right] \leq 110.6 \cdot RL \sqrt{T \cdot d^{1+\frac{3}{\sqrt{d}}} - \frac{1}{p}} + 5.9 \cdot \sqrt{R} (\sigma + L) T^{\frac{3}{2}} \sqrt{b_q(d)d^{\frac{3}{2}} - \frac{1}{p}}.
\]

**D Definition of \( \ell_2 \)-randomized estimator**

In this section we recall the algorithm of Shamir [33]. Let \( \zeta^o \in \mathbb{R}^d \) be distributed uniformly on \( \partial B_d^d \). Instead of the gradient estimator that we introduce in Algorithm 1, at a each step \( t \geq 1 \), Shamir [33] uses

\[
g_t^o = \frac{d}{2h} (y^{o'}_t - y_t') \zeta^o_t,
\]

where \( y^{o'}_t = f_t(x_t + h_t \zeta^o_t) \), \( y_t' = f_t(x_t - h_t \zeta^o_t) \), and \( \zeta^o_t \)'s are independent random variables with the same distribution as \( \zeta^o \).