Appendix
Fugue: Slow-Worker-Agnostic Distributed Learning for Big Models on Big Data

This appendix contains proof details for the main paper, as well as some experimental details. In this document, some of the theorems and equations refer to the main paper; we put "(main paper)" behind these references to avoid confusion.

A Parameter Tuning for Experiments

Model	TM	Dictionary Learning	MMND
Fugue's η_0	0.01	0.01	0.01
Fugue's $\eta_{'}$	0.01	0.01	0.05
BarrieredFugue's η_0	0.01	0.01	0.05
GraphLab's η_0	0.05	0.001	0.1
GraphLab's step_dec	0.9	0.9	0.9
PSGD's η_0	0.005	0.01	0.1

Table 1: Final tuned parameter values for Fugue, BarrieredFugue, GraphLab and PSGD. All the methods are tuned to perform optimally. η_0 is the initial step size, where η is defined in equation 4. λ is the Dictionary Learning ℓ_1 penalty defined in equation 2. $\eta^{'}$ is parameter that modifies the learning rate when extra updates are executed while waiting for slow workers. step_dec is a parameter for decreasing learning rate for GraphLab used in their collaborative filtering library.

Table 1 shows the final parameter values for each problem, after they have been tuned optimally for each method. $\eta^{'}$ controls the learning rate in additional updates by the worker when it is waiting for slower ones to finish. This makes it satisfy the condition in equation 11. The modified learning rate $\eta_t^{'} = \eta_t * (\eta^{'})^{x+1}$ where x is the number of extra update iterations the worker has perfomed while waiting in this sub-epoch. One iteration here means updating all the data points in the curent sub-epoch once. η_t is the sub-epoch's original learning rate without any extra updates similar to BarrieredFugue. The learning rate for epoch t is given by $\eta_t = \frac{\eta_0}{t+1}$ for Fugue and BarrieredFugue.

B Convergence Proof: Theorem 1 (main paper)

From equation 4 (main paper) we have

$$\begin{aligned}
\psi^{(t+1)} &= \psi^{(t)} - \eta_t \delta L^{(t)}(V^{(t)}, \psi^{(t)}) \\
&= \psi^{(t)} - \eta_t \nabla \mathcal{L}(\psi^{(t)}) + \eta_t \left[\nabla \mathcal{L}(\psi^{(t)}) - \delta L^{(t)}(V^{(t)}, \psi^{(t)}) \right] \\
&= \psi^{(t)} - \eta_t \nabla \mathcal{L}(\psi^{(t)}) + \eta_t \varepsilon_t \qquad (1)
\end{aligned}$$

Using n_i and N_w as defined in equation 6 (main paper)

$$\psi^{(t+(\sum_1^w n_i)m)} = \psi^{(t)} + \sum_{i=t}^{t+m(\sum_1^w n_i)} -\eta_i \nabla \mathcal{L}(\psi^{(i)}) + \sum_{i=t}^{t+m(\sum_1^w n_i)} \eta_i \varepsilon_i$$

$$\implies \psi^{(t+mN_w)} = \psi^{(t)} + \sum_{i=t}^{t+mN_w} -\eta_i \nabla \mathcal{L}(\psi^{(i)}) + \sum_{i=t}^{t+mN_w} \eta_i \varepsilon_i$$

$$\text{assuming} \sum_1^w n_i = N_w$$

$$\implies \psi^{(t+mN_w)} = \psi^{(t)} + \sum_{i=t}^{t+mN_w} -\eta_i \nabla \mathcal{L}(\psi^{(i)}) + M_{mN_w} \tag{2}$$

where $M_{mN_w} = \sum_{i=t}^{t+mN_w} \eta_i \varepsilon_i$ is a martingale sequence since it is a sum of martingale difference sequence. mN_w captures the m whole sub-epochs of work done as a whole by all the workers combined. From Doobs martingale inequality (Friedman, 1975, ch. 1, Thm 3.8)

$$P\left(\sup_{t+mN_w \geq r \geq t} |M_r| \geq c\right) \leq \frac{E\left[\left(\sum_{i=t}^{t+mN_w} \eta_i \varepsilon_i\right)^2\right]}{c^2} \tag{3}$$

where $M_r = \sum_{i=t}^{r} \eta_i \varepsilon_i$. Lets look at the RHS of equation 3 above:

$$E\left[\left(\sum_{i=t}^{t+mN_w} \eta_i \varepsilon_i\right)^2\right] = E\left[\sum_{i=1}^{mN_w} (\eta_i \varepsilon_i)^2\right]$$

$$(\text{equation 8 (main paper)} \implies E[\varepsilon_i \varepsilon_j] = 0 \text{ if } i \neq j)$$

$$= \sum_{i=1}^{mN_w} \eta_i^2 E[\varepsilon_i^2] \leq \sum_{i=1}^{mN_w} \eta_i^2 D \to 0$$

$$\text{where } E[\varepsilon_i^2] < D \; \forall i \text{ and assuming} \sum \eta_i^2 < \infty$$

$$\lim_{t \to \infty} \implies P\left(\sup_{i \geq t} |M_i| \geq c\right) = 0 \;\; as \;\; t \to \infty \tag{4}$$

From equation 4 we have

$$\psi^{(t+mN_w)} = \psi^{(t)} + \sum_{i=t}^{t+mN_w} -\eta_i \nabla \mathcal{L}(\psi^{(i)})$$

asymptotically.
Note that we do a theoretical analysis of the algorithm without projection steps. Extending the proof to include projection can be done by using Arzela-Ascoli theorem and the limits of converging sub-sequence of our algorithm's SGD updates (Kushner and Yin, 2003). ■

C Intra sub-epoch variance

$$\psi^{(t+1)} = \psi^{(t)} - \delta\psi^{(t)}(V^{(t)}, \psi^{(t)}) \tag{5}$$

$$\text{where } \delta\psi^{(t)}(V^{(t)}, \psi^{(t)}) = \eta^{(t)} \delta L^{(t)}(V^{(t)}, \psi^{(t)}) \Rightarrow \psi^{(t+1)} = \psi^t - \eta_t \delta L^t(V^t, \psi^t)$$

Summing equation 5 over n_i, the number of points updated in block i of a sub-epoch

$$\psi^{t+n_i} = \psi^t - \sum_{i=1}^{n_i} \eta_{t+i} \delta L^{t+i}(V^{t+i}, \psi^{t+i}) \tag{6}$$

As defined earlier, V denotes the joint potential for all the n_i points encountered in block i. The equation 7 (main paper) for V is

$$
\begin{aligned}
&p(\psi^{(t+n_i)}|\psi^t)d\psi^{(t+n_i)} = p(V(\psi^{(t+n_i)},\psi^t))dV \\
&\Rightarrow p(\psi^{(t+n_i)})d\psi^{(t+n_i)} = \int_{\psi^t} p(\psi^{(t+n_i)}|\psi^t)p(\psi^t)d\psi^t d\psi^{(t+n_i)} = \int_{\psi^t} p(V(\psi^{(t+n_i)},\psi^t))dV p(\psi^t)d\psi^t
\end{aligned} \tag{7}
$$

Lemma 1 *Let $u(\psi^{(t+n_i)})$ be a function of $\psi^{(t+n_i)}$ then*

$$\mathbb{E}^{\psi^{(t+n_i)}}[u(\psi^{(t+n_i)})] = \mathbb{E}^{\psi^t}[\mathbb{E}^V[u(\psi^{(t+n_i)})]]$$

Proof. From equation 7

$$
\begin{aligned}
\mathbb{E}^{\psi^{(t+n_i)}}[u(\psi^{(t+n_i)})] &= \int_{\psi^{(t+n_i)}} u(\psi^{(t+n_i)})p(\psi^{(t+n_i)})d\psi^{(t+n_i)} \\
&= \int_{\psi^{t+i}} u(\psi^{(t+n_i)})P(\psi^{(t+n_i)})d\psi^{(t+n_i)} \\
&= \int_V \int_{\psi^t} u(\psi^{(t+n_i)})P(V(\psi^{(t+n_i)},\psi))dV P(\psi^t)d\psi^t \\
&= \mathbb{E}^{\psi^t}[\mathbb{E}^V[u(\psi^{(t+n_i)})]]
\end{aligned}
$$

■

Lemma 2

$$\mathbb{E}^V[\delta L^{t+i}(v^{t+i},\psi^{t+i})] = \frac{d\mathbb{E}^V[L^{t+i}(v^{t+i},\psi^{t+i})]}{d\psi^{t+i}}$$

Proof. Due to randomness in picking the point to be updated in iteration $t+i$ We have

$$
\begin{aligned}
&\mathbb{E}^V[L^{t+i}(v^{t+i},\psi^{t+i})] = \int L(y,\psi^{t+i})dy \\
\Rightarrow &\frac{d\mathbb{E}^V[L^{t+i}(v^{t+i},\psi^{t+i})]}{d\psi^{t+i}} = \mathbb{E}^V[\frac{dL^{t+i}(v^{t+i},\psi^{t+i})}{d\psi^{t+i}}] = \mathbb{E}^V[\delta L^{t+i}(v^{t+i},\psi^{t+i})]
\end{aligned}
$$

■

Lemma 3

$$\mathbb{E}^V[L^{t+i}(v^{t+i},\psi^{t+i})] = \mathbb{E}^{v^{t+i}}[L^{t+1}(v^{t+i},\psi^{t+i})]$$

Proof.
Using the definition of V^{t+i} in equation 7 (main paper), the fact that V is a joint variable of each V^{t+i} and an any iteration $t+i$ the chance of picking any data point is completely random and indpendent of any other iteration.

$$\mathbb{E}^V[L^{t+i}(v^{t+i},\psi^{t+i})] = \mathbb{E}^{v^{t+i}}[L^{t+1}(v^{t+i},\psi^{t+i})]$$

■

Lemma 4

$$\mathbb{E}^V[\frac{dL^{t+i}(v^{t+i},\psi^{t+i})}{d\psi^{t+i}}\frac{dL^{t+j}(v^{t+j},\psi^{t+i})}{d\psi^{t+j}}] = \frac{d\mathbb{E}^{v^{t+i}}[L^{t+i}(v^{t+i},\psi^{t+i})]}{d\psi^{t+i}}\frac{d\mathbb{E}^{v^{t+i}}[L^{t+i}(v^{t+i},\psi^{t+i})]}{d\psi^{t+i}}$$

Proof. Two different data points picked at iteration $(t+i)$ and $(t+j)$ are independent of each other. Using this fact and the definition of potetntial function V in equation 7

$$\mathbb{E}^V[\frac{dL^{t+i}(v^{t+i},\psi^{t+i})}{d\psi^{t+i}}\frac{dL^{t+j}(v^{t+j},\psi^{t+i})}{d\psi^{t+j}}] = \mathbb{E}^V[\frac{dL^{t+i}(v^{t+i},\psi^{t+i})]}{d\psi^{t+i}}]\mathbb{E}^V[\frac{dL^{t+j}(v^{t+j},\psi^{t+i})}{d\psi^{t+j}}]$$
$$\left(\text{using lemma 2}\right) = \frac{d\mathbb{E}^V[L^{t+i}(v^{t+i},\psi^{t+i})]}{d\psi^{t+i}}\frac{d\mathbb{E}^V[L^{t+j}(v^{t+j},\psi^{t+j})]}{d\psi^{t+j}}$$
$$\left(\text{using lemma 3}\right) = \frac{d\mathbb{E}^{v^{t+i}}[L^{t+i}(v^{t+i},\psi^{t+i})]}{d\psi^{t+i}}\frac{d\mathbb{E}^{v^{t+j}}[L^{t+j}(v^{t+j},\psi^{t+j})]}{d\psi^{t+j}}$$

■

Theorem 1 *We define ψ_* as the global optima and Ω_0 as the hessian of the loss at ψ_* i.e. $\Omega_0 = \frac{d^2\mathbb{E}[L(\psi_*)]}{d\psi_*^2}$ (assuming that ψ is univariate) then*

$$\frac{d\mathbb{E}^{v^{t+i}}[L^{t+i}(v^{t+i},\psi^{t+i})]}{d\psi^{t+i}} = \Omega_0(\psi_t - \psi_* + \delta_i) + \mathcal{O}(\rho_t^2)$$

where $\mathcal{O}(\rho_t^2) = \mathcal{O}(|\psi_{t+i} - \psi_*|^2)$ with the assumption that $\mathcal{O}(\rho_{t+i})$ is small $\forall i \geq 0$ and $\delta_i = \psi_{t+i} - \psi_t$.

Proof. Lets define $\phi(\psi_{t+i}) = \mathbb{E}^{v^{t+i}}[L^{t+i}(v^{t+i},\psi^{t+i})]$ Using Taylor's theorem and expanding around ψ_*

$$\phi(\psi^{t+i}) = \phi(\psi_*) + \frac{d\phi(\psi_*)}{d\psi_*}(\psi^{t+i} - \psi_*) + \frac{(\psi^{t+1} - \psi_*)^2}{2}\frac{d^2\phi(\psi_*)}{d\psi_*^2} + \mathcal{O}((\psi^{t+i} - \psi_*)^3)$$
$$= \phi(\psi_*) + \frac{(\psi^{t+i} - \psi_*)^2}{2}\frac{d^2\phi(\psi_*)}{d\psi_*^2} + \mathcal{O}((\psi^{t+i} - \psi_*)^3)\left(\text{as } \frac{d\phi(\psi_*)}{d\psi_*} = 0 \text{ at optima}\right)$$
$$\Rightarrow \frac{d\phi(\psi^{t+i})}{d\psi^{t+i}} = (\psi^{t+i} - \psi_*)\frac{d^2\phi(\psi_*)}{d\psi_*^2} + \mathcal{O}((\psi^{t+i} - \psi_*)^2)$$
$$\Rightarrow \frac{d\mathbb{E}^{v^{t+i}}[L^{t+i}(v^{t+i},\psi^{t+i})]}{d\psi^{t+i}} = \Omega_0(\psi^t - \psi_* + \delta_i) + \mathcal{O}(\rho_t^2)\left(\text{with the assumption that } \mathcal{O}(\rho_t) \text{ is small we have } \mathcal{O}(\rho_{t+i}^2) = \mathcal{O}(\rho_t^2)\right)$$

■

Theorem 2 *With ψ_* as defined in theorem 1 and assuming that ψ is univariate we have*

$$\mathbb{E}^{v^{t+i}}[(\frac{dL^{t+i}(v^{t+i},\psi^{t+i})}{d\psi^{t+i}})^2] = \Omega_1 + \mathcal{O}(\mathbb{E}[\mathcal{O}(\rho_t)]) + \mathcal{O}(\rho_t^2)$$

where $\mathcal{O}(\rho_t^2)$ and δ_i are as defined in theorem 1 and $\Omega_1 = \mathbb{E}^{v^{t+i}}[(\frac{dL^{t+i}(v^{t+i},\psi_*)}{d\psi_*})^2]$

Proof. Expanding $L^{t+i}(v^{t+i}, \psi^{t+i})$ around ψ_* using Taylor's theorem

$$
\begin{aligned}
L^{t+i}(v^{t+i}, \psi^{t+i}) &= L^{t+i}(v^{t+i}, \psi_*) + \frac{dL^{t+i}(v^{t+i}, \psi_*)}{d\psi_*}(\psi^{t+i} - \psi_*) \\
&+\frac{1}{2}\frac{d^2L^{t+i}(v^{t+i}, \psi_*)}{d\psi_*^2}(\psi^{t+i} - \psi_*)^2 + \mathcal{O}((\psi^{t+i} - \psi_*)^3) \\
\Rightarrow \frac{dL^{t+i}(v^{t+i}, \psi^{t+i})}{d\psi^{t+i}} &= \frac{dL^{t+i}(v^{t+i}, \psi_*)}{d\psi_*} + \frac{d^2L^{t+i}(v^{t+i}, \psi_*)}{d\psi_*^2}(\psi^{t+i} - \psi_*) + \mathcal{O}((\psi^{t+i} - \psi_*)^2) \\
\Rightarrow \mathbb{E}^{v^{t+i}}[(\frac{dL^{t+i}(v^{t+i}, \psi^{t+i})}{d\psi^{t+i}})^2] &= \mathbb{E}^{v^{t+i}}[(\frac{dL^{t+i}(v^{t+i}, \psi_*)}{d\psi_*})^2 \\
+2\frac{dL^{t+i}(v^{t+i}, \psi_*)}{d\psi_*}\frac{d^2L^{t+i}(v^{t+i}, \psi_*)}{d\psi_*^2}&(\psi^{t+i} - \psi_*) + \mathcal{O}((\psi^{t+i} - \psi_*)^2)] \\
\Rightarrow \mathbb{E}^{v^{t+i}}[(\frac{dL^{t+i}(v^{t+i}, \psi^{t+i})}{d\psi^{t+i}})^2] &= \Omega_1 + \mathcal{O}(\mathbb{E}[(\psi_{t+i} - \psi_*]) + \mathcal{O}(\rho_t^2) \\
&= \Omega_1 + \mathcal{O}(\mathbb{E}[\mathcal{O}(\rho_t)]) + \mathcal{O}(\rho_t^2)
\end{aligned}
$$

■

C.1 Within block variance bound

Theorem 3 *The variance of the parameter ψ at the end of a sub-epoch S in block S_i which updated n_i points as defined in equation 6 (main paper) is*

$$
\begin{aligned}
Var(\psi^{t+n_i}) = &Var(\psi^t) - 2\eta_t n_i \Omega_0 (Var(\psi^t)) - 2\eta_t n_i \Omega_0 CoVar(\psi_t, \bar{\delta}_t) + \eta_t^2 n_i \Omega_1 \\
&+ \underbrace{\mathcal{O}(\eta_t^2\rho_t) + \mathcal{O}(\eta_t\rho_t^2) + \mathcal{O}(\eta_t^3) + \mathcal{O}(\eta_t^2\rho_t^2)}_{\Delta_t}
\end{aligned}
$$

Constants Ω_0 and Ω_1 are defined in theorems 1 and theorems 2 respectively.

Proof. We start with analysing $\mathbb{E}^V[u(\psi^{(t+ni)})]$ term from lemma 1

$$
\begin{aligned}
\mathbb{E}^V[u(\psi^{(t+ni)})] &= \mathbb{E}^V[u(\psi^t + \underbrace{(-\sum_{i=1}^{n_i}\eta_{t+i}\delta L^{t+i}(v^{t+i},\psi^{t+i}))}_{\nabla})] \\
&= \mathbb{E}^V[u(\psi^t) - \frac{du(\psi^t)}{d\psi^t}\nabla + \frac{1}{2}\frac{du^2(\psi^t)}{d(\psi^t)^2}\nabla^2 + \mathcal{O}(\eta_t^3)] \\
&= u(\psi^t) - \eta_t\frac{du(\psi^t)}{d\psi^t}\mathbb{E}^V[\sum_{i=1}^{n_i}\delta L^{t+i}(v^{t+i},\psi^{t+i})] + \eta_t^2\frac{1}{2}\frac{du^2(\psi^t)}{d(\psi^t)^2}\mathbb{E}^V[(\sum_{i=1}^{n_i}\delta L^{t+i}(v^{t+i},\psi^{t+i}))^2] \\
&+ \mathcal{O}(\eta_t^3) \qquad \left(\text{since } \eta_t = \eta_{t+i} \text{ within a block and expanding } \nabla\right) \\
&= u(\psi^t) - \eta_t\frac{du(\psi^t)}{d\psi^t}\sum_{i=1}^{n_i}\frac{d\mathbb{E}^V[L^{t+i}(v^{t+i},\psi^{t+i})]}{d\psi^{t+i}} + \eta_t^2\frac{1}{2}\frac{du^2(\psi^t)}{d(\psi^t)^2}\mathbb{E}^V[(\sum_{i=1}^{n_i}\frac{dL^{t+i}(v^{t+i},\psi^{t+i})}{d\psi^{t+i}})^2] \\
&+ \mathcal{O}(\eta_t^3) \\
&= u(\psi^t) - \eta_t\frac{du(\psi^t)}{d\psi^t}(\sum_{i=1}^{n_i}\frac{d\mathbb{E}^{v^{t+i}}[L^{t+i}(v^{t+i},\psi^{t+i})]}{d\psi^{t+i}}) \\
&\left(\text{using Lemma 3}\right) \\
&+ \eta_t^2\frac{1}{2}\frac{du^2(\psi^t)}{d(\psi^t)^2}\left[\mathbb{E}^V[\sum_{i=1}^{n_i}(\frac{dL^{t+i}(v^{t+i},\psi^{t+i})}{d\psi^{t+i}})^2] + \mathbb{E}^V[\sum_{i\neq j}\frac{dL^{t+i}(v^{t+i},\psi^{t+i})}{d\psi^{t+i}}\frac{dL^{t+j}(v^{t+j},\psi^{t+j})}{d\psi^{t+j}}]\right] \\
&+ \mathcal{O}(\eta_t^3) \\
&= u(\psi^t) - \eta_t\frac{du(\psi^t)}{d\psi^t}(\sum_{i=1}^{n_i}\frac{d\mathbb{E}^{v^{t+i}}[L^{t+i}(v^{t+i},\psi^{t+i})]}{d\psi^{t+i}}) + \eta_t^2\frac{1}{2}\frac{du^2(\psi^t)}{d(\psi^t)^2}\left[\sum_{i=1}^{n_i}\mathbb{E}^{v^{t+i}}[(\frac{dL^{t+i}(v^{t+i},\psi^{t+i})}{d\psi^{t+i}})^2]\right. \\
&\left.+ (\sum_{i\neq j}\frac{d\mathbb{E}^{v^{t+i}}[L^{t+i}(v^{t+i},\psi^{t+i})]}{d\psi^{t+i}}\frac{d\mathbb{E}^{v^{t+j}}[L^{t+j}(v^{t+j},\psi^{t+j})]}{d\psi^{t+j}})\right] + \mathcal{O}(\eta_t^3) \\
&\left(\text{using Lemma 4}\right)
\end{aligned} \tag{8}
$$

From equation 8 and lemma 1

$$
\begin{aligned}
\mathbb{E}^{\psi^{(t+n_i)}}[u(\psi^{(t+n_i)})] = \mathbb{E}^{\psi^{(t)}}\Bigg[&u(\psi^t) - \eta_t\frac{du(\psi^t)}{d\psi^t}(\sum_{i=1}^{n_i}\frac{d\mathbb{E}^{v^{t+i}}[L^{t+i}(v^{t+i},\psi^{t+i})]}{d\psi^{t+i}}) \\
&+ \eta_t^2\frac{1}{2}\frac{du^2(\psi^t)}{d(\psi^t)^2}\left[\sum_{i=1}^{n_i}\mathbb{E}^{v^{t+i}}[(\frac{dL^{t+i}(v^{t+i},\psi^{t+i})}{d\psi^{t+i}})^2]\right. \\
&\left.+ (\sum_{i\neq j}\frac{d\mathbb{E}^{v^{t+i}}[L^{t+i}(v^{t+i},\psi^{t+i})]}{d\psi^{t+i}}\frac{d\mathbb{E}^{v^{t+j}}[L^{t+j}(v^{t+j},\psi^{t+j})]}{d\psi^{t+j}})\right]\Bigg] + \mathcal{O}(\eta_t^3)
\end{aligned} \tag{9}
$$

From equation above the variance of ψ^{t+n_i} is

$$
\begin{aligned}
&Var(\psi^{t+n_i}) = \mathbb{E}^{\psi^{(t+n_i)}}[(\psi^{(t+n_i)})^2] - \Big(\mathbb{E}^{\psi^{(t+n_i)}}[\psi^{(t+n_i)}]\Big)^2 \\
&= \mathbb{E}^{\psi^t}[(\psi^t)^2] - \eta_t n_i \mathbb{E}^{\psi^t}[2\psi^t(\Omega_0(\psi^t - \psi_* + \bar{\delta}_t) + \mathcal{O}(\rho_t^2))] \\
&\left(\text{using theorem 1 and defining } \bar{\delta}_t = \frac{\sum_{i=1}^{n_i}\delta_i}{n_i}\right) \\
&+ \eta_t^2 \frac{1}{2}\mathbb{E}^{\psi^t}[2\{n_i(\Omega_1 + \mathcal{O}(\mathbb{E}[\rho_t]) + \mathcal{O}(\rho_t^2)) \\
&+ \sum_{i\neq j}(\Omega_0(\psi^{t+i} - \psi_*) + \mathcal{O}(\rho_t^2))(\Omega_0(\psi^{t+j} - \psi_*) + \mathcal{O}(\rho_t^2))\}] \\
&- \Big(\mathbb{E}^{\psi^t}[\psi^t] - \eta_t n_i \mathbb{E}^{\psi^t}[(\Omega_0(\psi^t - \psi_* + \bar{\delta}_t) + \mathcal{O}(|\psi^t - \psi_*|^2))]\Big)^2 \\
&= \mathbb{E}^{\psi^t}[(\psi^t)^2] - 2\Omega_0\eta_t n_i \mathbb{E}^{\psi^t}[(\psi^t)^2] + 2\Omega_0\eta_t n_i \psi_* \mathbb{E}^{\psi^t}[\psi^t] - 2\Omega_0\eta_t n_i \mathbb{E}^{\psi^t}[\psi^t\bar{\delta}_t] - \mathcal{O}(\eta_t\rho_t^2) \\
&+ \eta_t^2 n_i \Omega_1 + \mathcal{O}(\eta_t^2\rho_t) + \mathcal{O}(\eta_t^2\rho_t^2) + \mathcal{O}(\eta_t^2\rho_t^3) + \mathcal{O}(\eta_t^2\rho_t^4) \\
&- \Big(\mathbb{E}^{\psi^t}[\psi^t]\Big)^2 + 2n_i\eta_t\mathbb{E}^{\psi^t}[\psi^t](\mathbb{E}^{\psi^t}[\Omega_0\psi^t] - \Omega_0\psi_* + \mathbb{E}^{\psi^t}[\Omega_0\bar{\delta}_t] + \mathcal{O}(\rho_t^2)) - \mathcal{O}(\eta_t^2\rho_t^2) + \mathcal{O}(\eta_t^3) \\
&= Var(\psi^t) - 2\eta_t n_i \Omega_0 (Var(\psi^t)) - 2\eta_t n_i \Omega_0 CoVar(\psi_t, \bar{\delta}_t) + \eta_t^2 n_i \Omega_1 \\
&+ \underbrace{\mathcal{O}(\eta_t^2\rho_t) + \mathcal{O}(\eta_t\rho_t^2) + \mathcal{O}(\eta_t^3) + \mathcal{O}(\eta_t^2\rho_t^2)}_{\Delta_t}
\end{aligned}
\tag{10}
$$

■