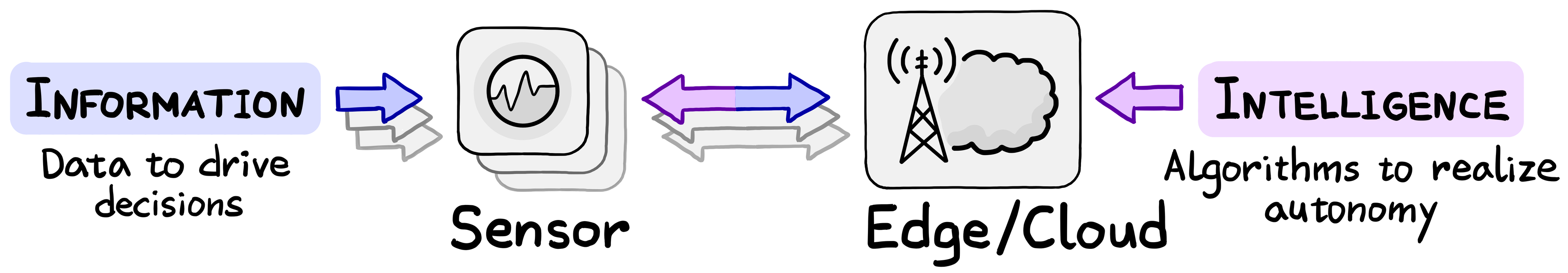
Randomized Edge-Assisted On-Sensor Information Selection for Bandwidth-Constrained Systems

Igor Burago
Marco Levorato
{iburago, levorato}@uci.edu
Department of Computer Science
University of California, Irvine

1. Intelligent IoT in Principle

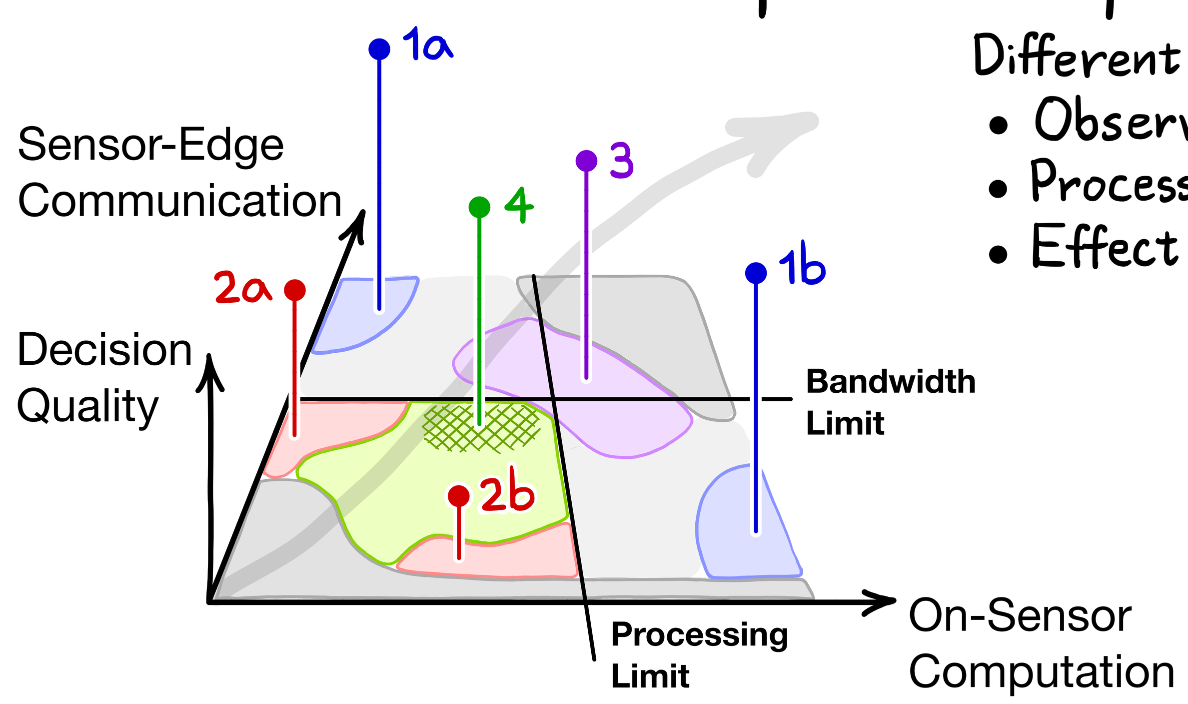


- Observes the environment
- Selectively informs the edge
- Adjusts its own operation

- Accumulates relevant observations
- Exercises high-level control
- Assists on-sensor decision-making

Goal: Realize the alliance between information completeness and best intelligence

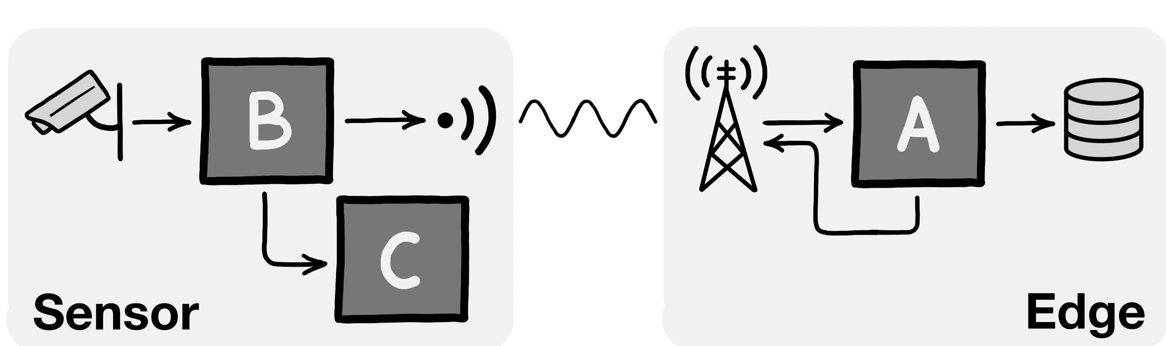
Schematic view into the space of computation-communication tradeoffs



Different strategies vary in the amount of:
- Observations transferred to the edge
- Processing allocated on the sensor
- Effect induced on the quality of decisions

1a Offload all decisions to edge
1b Offload all decisions to sensor
2a Sample observations to transmit
2b Transmit obvious positives
3 Transmit all in compressed form
4 Optimal adaptive compromise

2. Sensor-Edge Cooperation Model

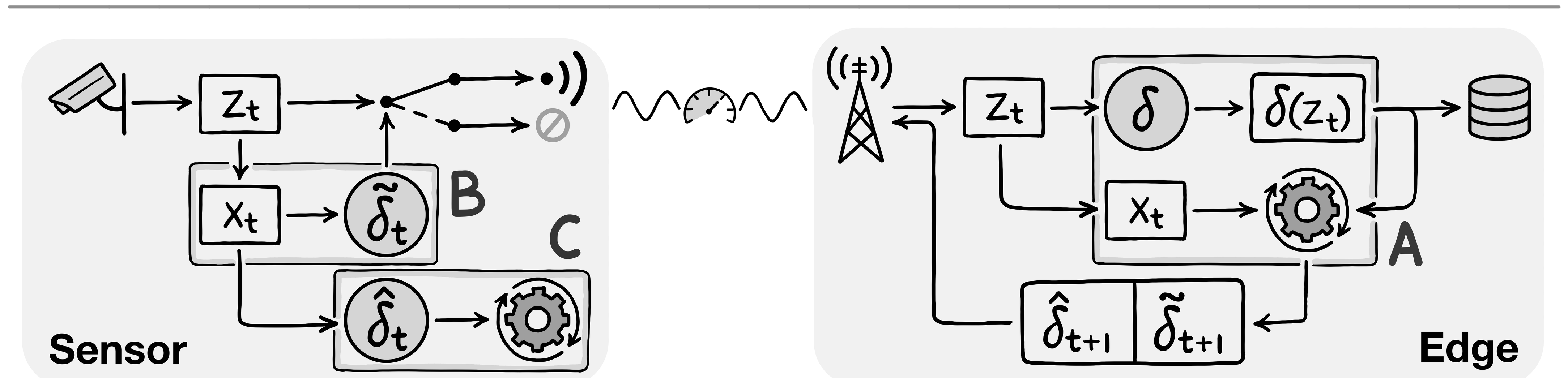


Model assumptions

- Observations fall into two classes: of interest to the system (positives) and not (negatives)
- Positives are to be collected at edge for further processing in the system
- Edge can run the optimal classifier δ; sensor can not
- B Sensor runs a low-complexity classifier supplied by edge
- A Edge, via supervision of its classifier, trains the sensor's
- C Sensor uses its classifier to spare bandwidth by dropping some negatives, and to make local decisions

Factors at play

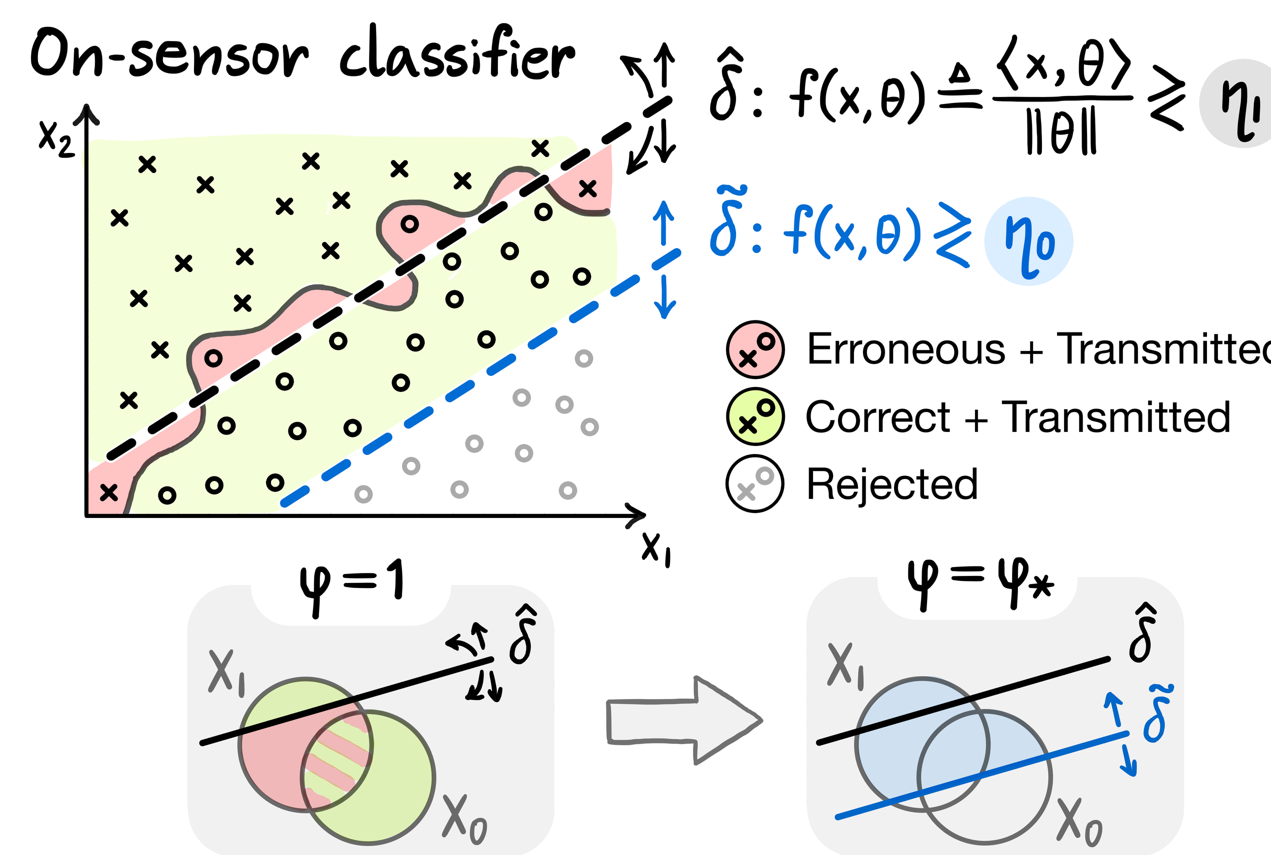
- CPU, battery, etc.
- Computation constraints
- Complexity of $\tilde{\delta}$ and $\hat{\delta}$
- Shared wireless channel
- Bandwidth constraint
- Transmission probability $\varphi \approx \varphi_*$
- Need to react to environment
- Cost of error in feature space
- Expected risk U to minimize



On-sensor classifier

$\delta:\ f(x,\theta) \triangleq \frac{\langle x,\theta\rangle}{\|\theta\|} \gtrless \eta_1$

$\tilde{\delta}:\ f(x,\theta) \gtrless \eta_0$



Problem: $U(\tau) \longrightarrow \min$, s.t. $W(\tau)=0$ $\Rightarrow$ $\frac{1}{\zeta}U(\tau)+W(\tau) \longrightarrow \min_\tau$

$U(\tau) \triangleq \mathbb{E}\left[\hat{\mathrm{I}}_{\eta_0}(u_\theta)\, \mathrm{I}_e(z,\tau)\, |r_0(u_\theta-\eta_0)\, r_1(u_\theta-\eta_1)|\right]$ — Expected penalty for errors in transmitted data

$W(\tau) \triangleq \frac{1}{2}(\varphi(\tau)-\varphi_*)^2$ — Constraint penalty

$\tau = \mathrm{vec}[\theta, \eta_0, \eta_1]$ — Parameters; $u_\theta \triangleq f(x,\theta)$ — Decision function; η_0, η_1 — Thresholds; $\varphi(\tau) \triangleq \mathbb{E}[\hat{\mathrm{I}}_{\eta_0}(u_\theta)]$ — Transmission probability

$\hat{\mathrm{I}}_{\eta_0}, \hat{\mathrm{I}}_{\eta_1}, \mathrm{I}_1 \in \{0,1\}$ indicate whether an observation z with features x is to be transmitted, classified as a positive, and is a true positive.
$\mathrm{I}_e(z,\tau) \triangleq (1-\mathrm{I}_1(z))\hat{\mathrm{I}}_{\eta_1}(u_\theta) + \mathrm{I}_1(z)(1-\hat{\mathrm{I}}_{\eta_1}(u_\theta))$ is the error indicator.

r_0, r_1 are loss functions based on "distances" $u_\theta-\eta_0$ and $u_\theta-\eta_1$. In the simplest case, $r_i(u_\theta-\eta_i) = u_\theta-\eta_i$.

3. Decision Rules Update

$$\theta_{t+1} = \theta_t - \gamma_{1,t}\hat{G}_\theta(\mathrm{z}_{t+1},\tau_t) - \gamma_{2,t}\hat{g}_\theta(\tilde{x}^{\varphi}_{t+1}, \tilde{x}^{\psi}_{t+1}, \tau_t, \varepsilon_t)$$
$$\eta_{0,t+1} = \eta_{0,t} - \gamma_{3,t}\hat{g}_{\eta_0}(\tilde{x}^{\varphi}_{t+1}, \tilde{x}^{\psi}_{t+1}, \tau_t, \varepsilon_t)$$
$$\eta_{1,t+1} = \max\{\eta_{0,t+1},\ \eta_{1,t} - \gamma_{1,t}\hat{G}_{\eta_1}(\mathrm{z}_{t+1},\tau_t)\} \quad (a)$$

$$\hat{G}_\theta(\{z_i\}_{i=1}^n,\tau) \triangleq \frac{1}{n}\sum_{i=1}^n\left[J(z_i,\tau)\nabla_\theta[r_0(u_i-\eta_0)r_1(u_i-\eta_1)]\right]$$
$$\hat{G}_{\eta_0}(\{z_i\}_{i=1}^n,\tau) \triangleq -\frac{1}{n}\sum_{i=1}^n\left[J(z_i,\tau)\frac{d}{du}r_0(u_i-\eta_0)\,r_1(u_i-\eta_1)\right]$$
$$\hat{G}_{\eta_1}(\{z_i\}_{i=1}^n,\tau) \triangleq -\frac{1}{n}\sum_{i=1}^n\left[J(z_i,\tau)\,r_0(u_i-\eta_0)\frac{d}{du}r_1(u_i-\eta_1)\right]$$
where $x_i = \chi(z_i)$, $u_i = f(x_i,\theta)$,
$$J(z,\tau) \triangleq \hat{\mathrm{I}}_{\eta_0}(f(\chi(z),\theta))\left(\hat{\mathrm{I}}_{\eta_1}(f(\chi(z),\theta)) - \mathrm{I}_1(z)\right) \quad (b)$$

$$\hat{g}_\theta(\tilde{x}^\varphi, \tilde{x}^\psi, \tau, \varepsilon) \triangleq (\hat{\varphi}(\tilde{x}^\varphi,\tau)-\varphi_*)\,\hat{\psi}_\theta(\tilde{x}^\psi,\tau,\varepsilon)$$
$$\hat{g}_{\eta_0}(\tilde{x}^\varphi, \tilde{x}^\psi, \tau, \varepsilon) \triangleq (\hat{\varphi}(\tilde{x}^\varphi,\tau)-\varphi_*)\,\hat{\psi}_{\eta}(\tilde{x}^\psi,\tau,\varepsilon)$$
$$\hat{\varphi}(\{x_i\}_{i=1}^n,\tau) \triangleq \frac{1}{n}\sum_{i=1}^n \hat{\mathrm{I}}_{\eta_0}(f(x_i,\theta))$$
$$\hat{\psi}_\theta(\{x_i\}_{i=1}^n,\tau,\varepsilon) \triangleq \frac{1}{\varepsilon n}\sum_{i=1}^n\left[\mathrm{I}_{S(\tau,\varepsilon)}(f(x_i,\theta))\nabla_\theta f(x_i,\theta)\right]$$
$$\hat{\psi}_\eta(\{x_i\}_{i=1}^n,\tau,\varepsilon) \triangleq -\hat{\psi}_0 = \mathrm{const} < 0 \quad (c)$$

4. Algorithm Specifics Overview

(a) Stochastic quasi-gradient algorithm

Updates are done in batches $\mathrm{z}_{t+1} = \{z_i\}$, with instrumental sub-batches $\tilde{x}^\varphi_{t+1}, \tilde{x}^\psi_{t+1}$ extracted from z_{t+1}. Threshold η_1 is clipped with η_0 for the bandwidth constraint to prevail over the classification risk.

(b) Quasi-gradient $\hat{G}$ of the risk U

$\mathbb{E}[\hat{G}(\mathrm{z},\tau)] = G(\tau)$, the generalized gradient such that $U(\tau_{t+1}) - U(\tau_t) \leqslant \langle \tau_{t+1}-\tau_t, G(\tau_t)\rangle + O(\|\tau_{t+1}-\tau_t\|^2)$.

(c) Quasi-gradient $\hat{g}$ of the constraint term W

- Evolution of the integral in the gradient of W: volume → surface → volume.

$\nabla_\tau W(\tau) = (\varphi(\tau)-\varphi_*)\psi(\tau)$, where

$$\psi(\tau) \triangleq \nabla_\tau \mathbb{E}[\hat{\mathrm{I}}_{\eta_0}(u_\theta)] = \int_{\{f(x,\theta)=\eta_0\}} \frac{\nabla_\tau[f(x,\theta)-\eta_0]}{\|\nabla_x f(x,\theta)\|}p(x)\,d\sigma$$
$$= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{S(\tau,\varepsilon)}\nabla_\tau[f(x,\theta)-\eta_0]\,p(x)\,dx$$

$S(\tau,\varepsilon) \triangleq \{x: \eta_0 < f(x,\theta) \leqslant \eta_0+\varepsilon\}$, $\mathrm{I}_{S(\tau,\varepsilon)}(u) \triangleq \mathbb{1}[u \in S(\tau,\varepsilon)]$

- Instrumental samples are used to make a stochastic estimator for the product of φ and ψ in $\nabla_\tau W(\tau)$:

$$\mathbb{E}[\hat{g}(\tilde{x}^\varphi, \tilde{x}^\psi, \tau, \varepsilon)] = \mathbb{E}[(\hat{\varphi}(x,\tau)-\varphi_*)\,\hat{\psi}(x,\tau,\varepsilon)] + O(\varepsilon)$$
$$= \mathbb{E}[\hat{\varphi}(\tilde{x}^\varphi,\tau)-\varphi_*]\,\mathbb{E}[\hat{\psi}(\tilde{x}^\psi,\tau,\varepsilon)] + O(\varepsilon)$$

- As in the above, the gradient of the integral leads to:

$$\hat{\psi}_\eta(\{x_i\}_{i=1}^n,\tau,\varepsilon) \triangleq -\frac{1}{\varepsilon n}\sum_{i=1}^n \mathrm{I}_{S(\tau,\varepsilon)}(f(x_i,\theta)) \leqslant 0$$

Yet, in the algorithm, we use a constant estimator for promptness in the updates of η_0, as higher adaptivity is of more value than accuracy for adjusting to the bandwidth constraint.

Theorem (Algorithm convergence) Let
- batches z_t be i.i.d.;
- sub-batches $\tilde{x}^\varphi_t$ and $\tilde{x}^\psi_t$ be mutually independent;
- $r_0(u)$, $r_1(u)$, and $f(x,\theta)$ be continuously differentiable;
- $\mathrm{sgn}(r_i(u)) = \mathrm{sgn}(u)$, $i\in\{0,1\}$; $r_i(u) \leqslant r_i(v)$ for all $0 \leqslant u < v$;
- p.d.f. $p(x)$ of observations in the feature space be continuous and have compact support;
- $\gamma_{1,t}/\gamma_{2,t}$ and ε_t be monotonically decreasing

$\sum \gamma_{i,t}^2 < \infty$, $i\in\{1,2,3\}$; $\sum \gamma_{j,t}\varepsilon_t < \infty$, $j\in\{1,2\}$;
$\sum \gamma_{3,t}\gamma_{1,t}/\gamma_{2,t} < \infty$, $\sum \gamma_{3,t} = \infty$, $\sum \gamma_{1,t}^2/\gamma_{2,t} = \infty$.

Then, for $\zeta_t = \gamma_{2,t}/\gamma_{1,t}$,
- $\frac{1}{\zeta_t}U(\tau_t) + W(\tau_t) \xrightarrow[t\to\infty]{a.s.} V_*$, $\mathbb{E}[V_*] < \infty$;
- $\lim_{t\to\infty} |\varphi(\tau_t) - \varphi_*| = 0$;
- $\lim_{t\to\infty} G_{\eta_1}(\tau_t) = 0$;
- $\lim_{t\to\infty} G_\theta(\tau_t) + \zeta_t\nabla_\theta W(\tau_t) = 0$.

2018-10-30 Pacific Grove, CA