Bias of fluid / mean field approximation

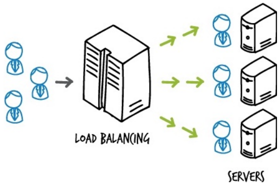
Poisson equation, averaging methods and two-timescale processes

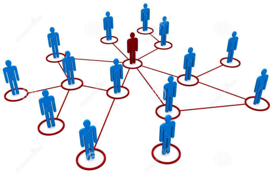
Nicolas Gast (Inria, Grenoble)

joint work with Sebastian Allmeier (Inria)

Séminaire Inria Paris, December 2022

Motivation: Studying interacting particle systems

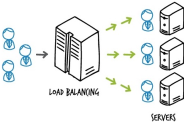




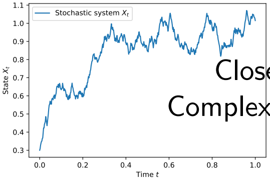
- Stochastic models are complex.

Fluid / mean field approximation simplifies the analysis

n servers



stochastic model

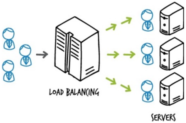


Close form?

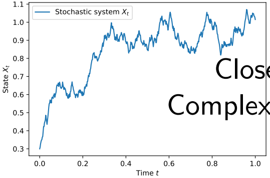
Complex to analyze

Fluid / mean field approximation simplifies the analysis

n servers

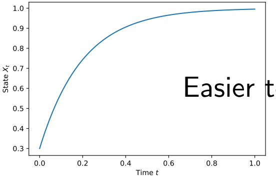


stochastic model



Close form?
Complex to analyze

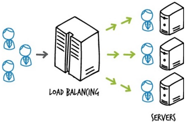
fluid



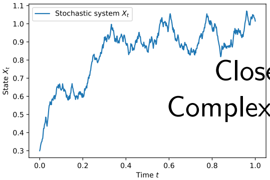
Easier to analyze

Fluid / mean field approximation simplifies the analysis

n servers



stochastic model

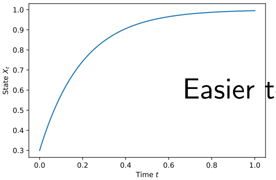


Close form?

Complex to analyze

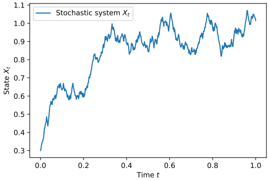
Question: What is the error made?

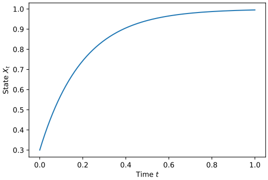
fluid



Easier to analyze

Fluid approximation is often justified by a law of large numbers

$\lim_{n\to\infty}$ (


) =


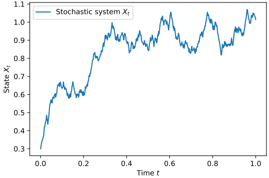
almost surely.

- Bound between X_t and $\phi_t(X_0)$ by using Gronwall's lemma.

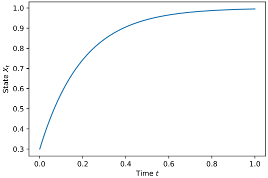
$$X_t - X_0 - \int_0^t f(X_s)ds \text{ is a martingale.}$$

Fluid approximation is often justified by a law of large numbers

$\lim_{n\to\infty}$ (



) =



almost surely.

- Bound between X_t and $\phi_t(X_0)$ by using Gronwall's lemma.

$$X_t - X_0 - \int_0^t f(X_s)ds \text{ is a martingale.}$$

- This gives a $O(1/\sqrt{n})$ convergence-rate.

In this talk: tools to provide sharp convergence results.

(Main) Related work

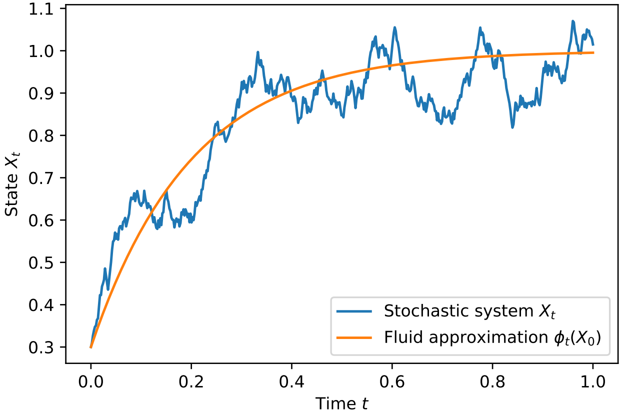
- Kurtz, 70s.
  - Fluid limits, diffusion limits (mostly transient regime)
- Stein's method:
  - Stein (1986)
  - Application to queueing: Braverman, Dai (2017–)
  - Application to mean-field models: Ying (2017).
- Refined mean field / Size expansions
  - Computational biology: Grima et al (2010s)
  - G. Van Houdt (2018), Allmeier G. (2021,2022).

Outline

Outline

1. Generators and Stein's method
2. Application to classical density dependent processes
3. Two time-scale processes
4. Conclusion

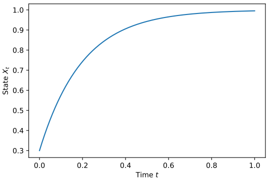
We compare a stochastic system and a fluid approximation



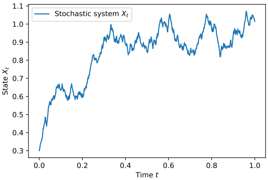
Important notations:

- Stochastic system $X_t \in \mathcal{X}$.
- Fluid approximation $\dot{x} = f(x)$. Solution starting from X_0 is $\phi_t(X_0)$.

To compare X_t and $\phi_t(X_0)$, we zoom on infinitesimal changes



$\Phi_t(X_0)$

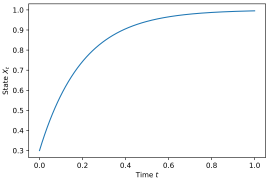


X_t

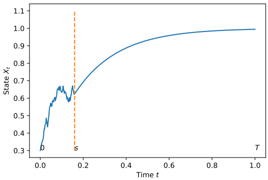
We want to compare:

$$\mathbb{E}[X_t] - \phi_t(X_0)$$

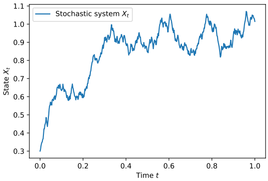
To compare X_t and $\phi_t(X_0)$, we zoom on infinitesimal changes



$\Phi_t(X_0)$



$\phi_{t-s}(X_s)$

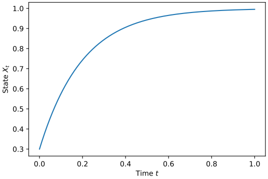


X_t

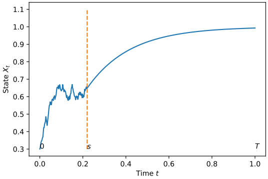
We want to compare:

$$\mathbb{E}\left[X_t\right] - \phi_t(X_0)$$

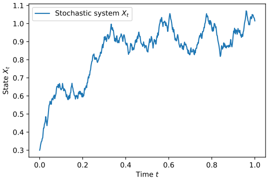
To compare X_t and $\phi_t(X_0)$, we zoom on infinitesimal changes



$\Phi_t(X_0)$



$\phi_{t-s}(X_s)$



X_t

We want to compare:

$$\mathbb{E}\left[X_t\right] - \phi_t(X_0) = \int_0^t \mathbb{E}\left[\frac{d}{ds}\phi_{t-s}(X_s)\right] ds$$

To study infinitesimal changes, we need generators

G^{sto} Generator of the stochastic system. For a test function $h : \mathcal{X} \to \mathbb{R}$:

$$G^{\mathrm{sto}} h(x) = \lim_{t \to 0} \frac{1}{t} \mathbb{E}\left[h(X_t) - h(X_0) \mid X_0 = x\right].$$

Example: if X_t is a Markov chain of generator K:

$$G^{\mathrm{sto}} h(x) = \sum_{y \in \mathcal{X}} K_{xy}(h(y) - h(x)).$$

To study infinitesimal changes, we need generators

G^{sto} Generator of the stochastic system. For a test function $h : \mathcal{X} \to \mathbb{R}$:

$$G^{\mathrm{sto}} h(x) = \lim_{t\to 0} \frac{1}{t} \mathbb{E}\left[h(X_t) - h(X_0) \mid X_0 = x\right].$$

G^{ODE} Generator of the ODE. For a test function $h : \mathcal{X} \to \mathbb{R}$:

$$G^{\mathrm{ODE}} h(x) = \lim_{t\to 0} \frac{1}{t} (h(\Phi_t(x)) - h(x))$$
$$= \nabla h(x) \cdot f(x).$$

Typically $f(x) = G^{\mathrm{sto}} I(x)$, where I is the identity function.

To study infinitesimal changes, we need generators

G^{sto} Generator of the stochastic system. For a test function $h : \mathcal{X} \to \mathbb{R}$:

$$G^{\mathrm{sto}} h(x) = \lim_{t \to 0} \frac{1}{t} \mathbb{E}\left[h(X_t) - h(X_0) \mid X_0 = x\right].$$

G^{ODE} Generator of the ODE. For a test function $h : \mathcal{X} \to \mathbb{R}$:

$$G^{\mathrm{ODE}} h(x) = \lim_{t \to 0} \frac{1}{t}(h(\Phi_t(x)) - h(x))$$

Using the generators, we can compare the two systems

$$\mathbb{E}\left[\frac{d}{ds}\phi_{t-s}(X_s)\right] ds$$

Using the generators, we can compare the two systems

$$
\begin{aligned}
\mathbb{E}\left[\frac{d}{ds}\phi_{t-s}(X_s)\right] ds &= \mathbb{E}\left[G^{\mathrm{sto}}\phi_{t-s}(X_s) - G^{\mathrm{ODE}}\phi_{t-s}(X_s)\right] \\
&= \mathbb{E}\left[(G^{\mathrm{sto}} - G^{\mathrm{ODE}})\phi_{t-s}(X_s)\right]
\end{aligned}
$$

Using the generators, we can compare the two systems

$$\mathbb{E}\left[\frac{d}{ds}\phi_{t-s}(X_s)\right] ds = \mathbb{E}\left[G^{\mathrm{sto}}\phi_{t-s}(X_s) - G^{\mathrm{ODE}}\phi_{t-s}(X_s)\right]$$
$$= \mathbb{E}\left[(G^{\mathrm{sto}} - G^{\mathrm{ODE}})\phi_{t-s}(X_s)\right]$$

In particular:

$$\mathbb{E}\left[X_t\right] - \phi_t(X_0) = \int_0^t \mathbb{E}\left[(G^{\mathrm{sto}} - G^{\mathrm{ODE}})\phi_{t-s}(X_s)\right] ds$$
$$= (G^{\mathrm{sto}} - G^{\mathrm{ODE}}) \int_0^t \mathbb{E}\left[\phi_{t-s}(X_s)\right] ds$$

$$\left(= (G^{\mathrm{ODE}} - G^{\mathrm{sto}}) \int_0^t \mathbb{E}\left[\phi_s(X_{t-s})\right] ds \right)$$

Taking the limit $t \to \infty$, we obtain Stein's method

This also works for $t = +\infty$ if X_t has a stationary distribution:

$$\mathbb{E}\left[X_\infty\right] - \phi_\infty(X_0) = \left(G^{\mathrm{ODE}} - G^{\mathrm{sto}}\right) \underbrace{\int_0^\infty \left(\mathbb{E}\left[\phi_s(X_\infty)\right] - \phi_\infty(X_0)\right) ds}_{\text{solution of a Poisson equation}}$$

(provided that the above make sense)

Zoom on the Poisson equation

Let G be a generator that has a stationary distribution π. For a function h, we denote by $\bar{h} = \sum_x h(x)\pi(x)$.

- A solution of the poisson equation is a function P_h such that

$$GP_h(x) = \bar{h} - h(x). \tag{1}$$

  - Solution to (1) is not unique in general.
  - Unique up to additive constant if the process is unichain.

Zoom on the Poisson equation

Let G be a generator that has a stationary distribution π. For a function h, we denote by $\bar{h} = \sum_x h(x)\pi(x)$.

- A solution of the poisson equation is a function P_h such that

$$GP_h(x) = \bar{h} - h(x). \qquad (1)$$

  - Solution to (1) is not unique in general.
  - Unique up to additive constant if the process is unichain.

- $P_h(x)$ represents how far x is from the stationary distribution. Indeed, if X_t is a system whose generator is G, a solution to this equation is:

$$P_h(x) = \int_0^\infty (h(X_t) - \bar{h})dt.$$

A solution of a Poisson equation is a bias

Link with MDPs and Markov reward process
Consider a Markov chain (X_t) and assume that you earn a reward h. Then:

$$\int_0^T \mathbb{E}\left[h(X_t)\right] dt = T\bar{h} + \underbrace{P_h(X_0)}_{\text{bias}} + o(1).$$

If the Markov chain has a transition kernel K, then P_h satisfies:

$$P_h(x) - \bar{h} = h(x) + \sum_y K_{xy} P_h(y).$$

Recap

For finite t:

$$\underbrace{\mathbb{E}[X_t]}_{\text{Stochastic system}} - \underbrace{\phi_t(X_0)}_{\text{deterministic approx.}} = (G^{\mathrm{sto}} - G^{\mathrm{ODE}}) \int_0^t \mathbb{E}[\phi_{t-s}(X_s)]\, ds$$

For $t = +\infty$, if $x^* = \phi_\infty(X_0)$ does not depend on X_0, we have:

$$\underbrace{\mathbb{E}[h(X_\infty)]}_{\text{Stochastic system}} - \underbrace{h(x^*)}_{\text{deterministic approx.}} = (G^{\mathrm{ODE}} - G^{\mathrm{sto}}) P_h(X_\infty).$$

Recap

For finite t:

$$\underbrace{\mathbb{E}\left[X_t\right]}_{\text{Stochastic system}} - \underbrace{\phi_t(X_0)}_{\text{deterministic approx.}} = (G^{\mathrm{sto}} - G^{\mathrm{ODE}}) \int_0^t \mathbb{E}\left[\phi_{t-s}(X_s)\right] ds$$

For $t = +\infty$, if $x^* = \phi_\infty(X_0)$ does not depend on X_0, we have:

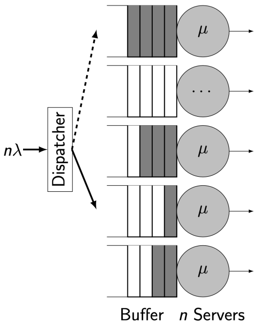
$$\underbrace{\mathbb{E}\left[h(X_\infty)\right]}_{\text{Stochastic system}} - \underbrace{h(x^*)}_{\text{deterministic approx.}} = (G^{\mathrm{ODE}} - G^{\mathrm{sto}}) P_h(X_\infty).$$

To prove that the sto $\approx$ deterministic, we prove that:

- for some $h \in \mathcal{H}$, $(G^{\mathrm{ODE}} - G^{\mathrm{sto}})h$ is small.
- $\int_0^t \mathbb{E}\left[\phi_{t-s}\right]$ or P_h belongs to this $\mathcal{H}$.

Outline

Classical Mean Field Setting



Example: Load-balancing

Mean Field Methodology:

- $X_s^{(n)}(t) = \frac{1}{n}\{\#$ objects in state s at $t\}$

Kurtz's density dependent population model:

$$X^{(n)} \to X^{(n)} + \frac{1}{n}\ell \qquad \text{at rate } n\beta_\ell(X)$$

Drift : $f(x) = \sum_\ell \ell\beta_\ell(x).$

Mean field approximaiton and result

Consider a density dependent population process in $\mathbb{R}^d$ and assume that $\beta_\ell(x)$ are bounded.

Theorem (G., Bortolussi, Tribastone 2019) If the drift is C^2, there exists an (easily computable) vector $V(t)$ such that for any finite time:

$$\mathbb{E}[X_t] = \phi_t(X_0) + \frac{1}{n}V(t) + O(\frac{1}{n^2}).$$

This holds uniformly in time if the ODE has a unique exponentially stable attractor.

$V(t)$ is the first-order expansion of the bias of the approximation.

The expansion is in general very accurate for small values of n

	Coupon	Supermarket	Pull/push
Simulation ($N = 10$)	1.530	2.804	2.304
Refined mean field ($N = 10$)	1.517	2.751	2.295
Mean field ($N = \infty$)	1.250	2.353	1.636

Table: Table from "A refined mean field approximation" (G. Van Houdt Sigmetrics 2018)

where

- mean field $= \Phi_t(x)$.
- Refined $=$ mean-field $+ V/n$.

Proof (1/2) Generator of a density dependent population process

The generator of the density dependent population process is:

$$G^{\mathrm{sto}}h(x) = \sum_{\ell} \left(h(x + \frac{1}{n}\ell) - h(x) \right) n\beta_{\ell}(x)$$

$$= \underbrace{\nabla h \cdot f(x)}_{\text{generator of the ODE, } G^{\mathrm{ODE}}} + \frac{1}{n} \underbrace{\nabla^2 h \cdot Q(x)}_{\text{Diffusion term}} + O(1/n^2).$$

Proof (1/2) Generator of a density dependent population process

The generator of the density dependent population process is:

$$
\begin{aligned}
G^{\mathrm{sto}}h(x) &= \sum_{\ell} \left(h(x + \frac{1}{n}\ell) - h(x) \right) n\beta_\ell(x) \\
&= \underbrace{\nabla h \cdot f(x)}_{\text{generator of the ODE, } G^{\mathrm{ODE}}} + \frac{1}{n} \underbrace{\nabla^2 h \cdot Q(x)}_{\text{Diffusion term}} + O(1/n^2).
\end{aligned}
$$

As a consequence

- $(G^{\mathrm{sto}} - G^{\mathrm{ODE}})h = O(1/n)$ if h is C^2. $\Rightarrow$ Set $\mathcal{H}$ of slide 15 is C^2.
- The hidden constant depends on $\left\|\nabla^2 h \cdot Q\right\|$. Studying this gives $V(t)$.

Proof (2/2) Consequence for the error of mean field model

- For finite-horizon, the function $h(x) = \int_0^t \phi_s(x) ds$ is C^2 if the drift function f is C^2.

- For infinite-horizon model, $h(x) = \int_0^\infty (\phi_s(x) - x^*) ds$ is C^2 if in addition x^* is an exponentially stable attractor.

The two functions belongs to "$\mathcal{H}$" $\Rightarrow$ Error $= O(1/n)$.

Some historical remarks

- Ying 2016: L_2 error is $O(1/\sqrt{n})$ for steady-state.
- G. 2017: Bias is $O(1/n)$.
- G. 2018, 2019: Expansion terms for the bias.

- G. Allmeier 2022: Extension to heterogeneous models.

Missing part: multi-scale models?

Outline

What is a two-timescale stochastic process?

Slow process $X \in \mathbb{R}^{d_x}$	Fast process $Y \in \{1 \dots d_y\}$
X jumps from x to $x + \frac{1}{n}\ell$ • at rate $n\beta_\ell(x, y)$.	Y jumps from y to y' • at rate $nK^{fast}_{y,y'}(x)$.

What is a two-timescale stochastic process?

Slow process $X \in \mathbb{R}^{d_x}$	Fast process $Y \in \{1 \dots d_y\}$
X jumps from x to $x + \frac{1}{n}\ell$ • at rate $n\beta_\ell(x, y)$.	Y jumps from y to y' • at rate $nK^{fast}_{y,y'}(x)$.

Two approximations:

Drift $\dot{X} \approx f(X, Y)$	$\mathbf{P}[Y(t) = y] \approx \pi_y(X(t))$

What is a two-timescale stochastic process?

Slow process $X \in \mathbb{R}^{d_x}$	Fast process $Y \in \{1 \dots d_y\}$
X jumps from x to $x + \frac{1}{n}\ell$ • at rate $n\beta_\ell(x, y)$.	Y jumps from y to y' • at rate $nK^{fast}_{y,y'}(x)$.

Two approximations:

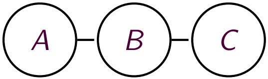
Drift $\dot{X} \approx f(X, Y)$	$\mathbf{P}[Y(t) = y] \approx \pi_y(X(t))$

$$\text{The ODE is } \dot{x} = \sum_y \pi_y(x) f(x, y)$$

(Averaging technique)

Example: CSMA with queues

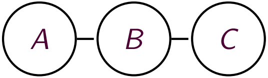
Model from Cecchi, Borst, Leeuwaarden 2015.



Interference graph, n nodes per class A, B or C.

Example: CSMA with queues

Model from Cecchi, Borst, Leeuwaarden 2015.



Interference graph, n nodes per class A, B or C.

"Slow process": X.

$X_{i,s}$ = proportion of nodes of class i with $\geq s$ messages

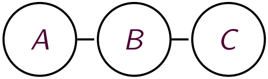
Arrival/departure:

$$X_{i,s} \mapsto X_{i,s} \pm \frac{1}{n}$$

Rate depends on Y.

Example: CSMA with queues

Model from Cecchi, Borst, Leeuwaarden 2015.



Interference graph, n nodes per class A, B or C.

"Slow process": X.

$X_{i,s}$ = proportion of nodes of class i with $\geq s$ messages

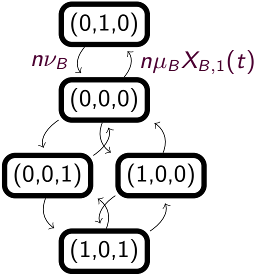
Arrival/departure:

$$X_{i,s} \mapsto X_{i,s} \pm \frac{1}{n}$$

Rate depends on Y.

"Fast process": Y.

$Y_i = 1$ if class i talks.



Accuracy results (Allmeier, G. 2022)

We consider a “generic” multiscale model as in Slide 24.

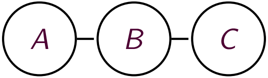
Theorem. Assume that $f(X, Y)$ and $K(X)$ are twice differentiable in X, and that $K^{fast}(X)$ is "unichain" for all X, then:

$$\mathbb{E}\left[X(t)\right] = \phi_t(X_0) + \frac{1}{n}V(t) + O(1/n^2).$$

This holds uniformly in time if the ODE has an exponentially stable attractor.

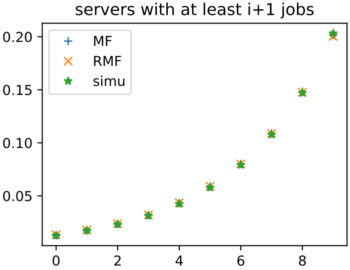
Again, the refined approximation is very accurate

Example with $n = 1$ node per class.



Jobs arrive at rate 1, activation rate $= 3$. Job duration is $1/3$.

Class A or C



Class B

Proof (1/3): comparison of generator is not sufficient!

Let $h : \mathcal{X} \to \mathbb{R}$ be a test function. We have:

$$G^{\mathrm{sto}}h(X,Y) = \sum_{\ell}(h(X+\frac{1}{n}\ell) - h(X))n\beta_\ell(X,Y)$$
$$= \nabla h \cdot f(X,Y) + O(1/n^2).$$
$$G^{\mathrm{ODE}}h(X,Y) = \nabla h \cdot \sum_{y}\pi_y(X)f(X,Y).$$

Proof (1/3): comparison of generator is not sufficient!

Let $h : \mathcal{X} \to \mathbb{R}$ be a test function. We have:

$$G^{\mathrm{sto}}h(X,Y) = \sum_{\ell}(h(X+\frac{1}{n}\ell) - h(X))n\beta_\ell(X,Y)$$
$$= \nabla h \cdot f(X,Y) + O(1/n^2).$$
$$G^{\mathrm{ODE}}h(X,Y) = \nabla h \cdot \sum_y \pi_y(X) f(X,Y).$$

Hence:

$$(G^{\mathrm{sto}} - G^{\mathrm{ODE}})h(X,Y) = \nabla h \cdot \underbrace{\left(f(X,Y) - \sum_y \pi_y(X) f(X,Y) \right)}_{\text{this was } = 0 \text{ for the singlescale model.}} + O(1/n)$$

Proof (2/3): Introducing the Poisson equation for the fast system

Let P_f^{fast} the solution of the Poisson equation for $K^{fast}(x)$ and f. Then:

$$f(X,Y) - \sum_y \pi_y(X) f(X,Y) = K^{fast}(x) P_f^{fast}(X,Y).$$

Lemma: if $K(x)$ is C^2 and unichain for all x, then P_f^{fast} is C^2.

Proof (3/3)

Let $h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a test function.

$$G^{\mathrm{sto}} h(X, Y) = nK^{fast} h(X, Y) + \nabla_x h \cdot f(X, Y) + O(1/n)$$

Hence, $K^{fast} = \frac{1}{n} G^{\mathrm{sto}}$ if h is C^1.

Proof (3/3)

Let $h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a test function.

$$G^{\mathrm{sto}} h(X, Y) = nK^{fast} h(X, Y) + \nabla_x h \cdot f(X, Y) + O(1/n)$$

Hence, $K^{fast} = \frac{1}{n} G^{\mathrm{sto}}$ if h is C^1.

Going back to the previous slides, we have:

$$(G^{\mathrm{sto}} - G^{\mathrm{ODE}}) h(X, Y) = K^{fast} \nabla h \cdot P_f^{fast}(X, Y) + O(1/n).$$

Hence, we are left with terms of the form

$$\begin{aligned}
\int_0^s K^{fast} P_f^{fast}(X_s, Y_s) ds &= \int_0^s \frac{1}{n} G^{\mathrm{sto}} P_f^{fast}(X_s, Y_s) ds \\
&= \frac{1}{n}(P_f^{fast}(X_t, Y_t) - P_f^{fast}(X_0, Y_0)) \\
&= O(1/n).
\end{aligned}$$

The last is because P_f^{fast} is C^1 when f and K are C^1.

Recap on multi-scale

This shows that the $O(1/n)$-expansion also holds for multiscale model provided that:

- Transitions are C^2 (as always).
- $K(x)$ is unichain for all x.

We can compute the expansion-term for $t = +\infty$.

Outline

Conclusion

Mean field or fluid approximations are widely used heuristic.

- They simplify the analysis of stochastic systems.

We question its validity / accuracy.

Conclusion

Mean field or fluid approximations are widely used heuristic.

- They simplify the analysis of stochastic systems.

We question its validity / accuracy.

- We characterize the bias for different models (smooth homogeneous, heterogeneous, multi-scale).
- To do so, we take correlations into account.
- Numerical library: `https://pypi.org/project/rmftool/`

Conclusion

Mean field or fluid approximations are widely used heuristic.

- They simplify the analysis of stochastic systems.

We question its validity / accuracy.

- We characterize the bias for different models (smooth homogeneous, heterogeneous, multi-scale).
- To do so, we take correlations into account.
- Numerical library: `https://pypi.org/project/rmftool/`

Many open questions: optimizaton (bandit problems), (sparse) geometric models, non-Markovian.

More slides and references: `http://polaris.imag.fr/nicolas.gast`

References

Results on which this talk is based:

- Mean Field and Refined Mean Field Approximations for Heterogeneous Systems: It Works! by Allmeier and Gast. SIGMETRICS 2022.
- A Refined Mean Field Approximation by Gast and Van Houdt. SIGMETRICS 2018 (best paper award)
- Size Expansions of Mean Field Approximation: Transient and Steady-State Analysis Gast, Bortolussi, Tribastone. Performance 2018.
- Two-scale: Bias and Refinement of Multiscale Mean Field Models. Allmeier, Gast, 2022 (arxiv).
  - CSMA model from CSMA networks in a many-sources regime: A mean-field approach. Cecchi, Borst, van Leeuwaarden, Whiting. Infocom 2016.

Paper cited as open problems:

- Pair-approximation: The Power of Two Choices on Graphs: the Pair-Approximation is Accurate by Gast. Mama 2015.
- Non-Markovian: Randomized Load Balancing with General Service Time Distributions by Bramson, Ly and Prabhakar. Sigmetrics 2010 and The PDE Method for the Analysis of Randomized Load Balancing Networks by Aghajani, Li, Ramanan.SIGMETRICS 2018