Mean Field Methods for Computer and Communication Systems

Jean-Yves Le Boudec
EPFL
July 2010
Contents

- Mean Field Interaction Model
- The Mean Field Limit
- Convergence to Mean Field
- The Decoupling Assumption
- Optimization
Mean Field Interaction Model: Common Assumptions

- Time is discrete or continuous

- $N$ objects
- Object $n$ has state $X_n(t)$
- $(X^N_1(t), ..., X^N_N(t))$ is Markov
  => $M^N(t)$ = occupancy measure process is also Markov

- Objects can be observed only through their state

- $N$ is large

Called “Mean Field Interaction Models” in the Performance Evaluation community

[McDonald(2007), Benaïm and Le Boudec(2008)]
Intensity $I(N)$

- $I(N) = \text{expected number of transitions per object per time unit}$

- The mean field limit occurs when we re-scale time by $I(N)$ i.e. we consider $X^N(t/I(N))$

- In discrete time
  - $I(N) = O(1)$: mean field limit is in discrete time
  - $I(N) = O(1/N)$: mean field limit is in continuous time
Example: 2-Step Malware

- Mobile nodes are either
  - `S` Susceptible
  - `D` Dormant
  - `A` Active

- Time is discrete

- Nodes meet pairwise (bluetooth)

- One interaction per time slot,
  \[ I(N) = 1/N; \] mean field limit is an ODE

- State space is finite
  \[ = \{S, A, D\}\]

- Occupancy measure is
  \[ M(t) = (S(t), D(t), A(t)) \] with
  \[ S(t) + D(t) + A(t) = 1 \]

- Possible interactions:

  1. Recovery
     - \[ D \rightarrow S \]
  2. Mutual upgrade
     - \[ D + D \rightarrow A + A \]
  3. Infection by active
     - \[ D + A \rightarrow A + A \]
  4. Recovery
     - \[ A \rightarrow S \]
  5. Recruitment by Dormant
     - \[ S + D \rightarrow D + D \]
     - Direct infection
     - \[ S \rightarrow D \]
  6. Direct infection
     - \[ S \rightarrow A \]

[Benâïm and Le Boudec(2008)]
Simulation Runs, N=1000 nodes

= D
State = A
State = S

Node 1

Node 2

Node 3

D(t)
Proportion of nodes
In state i=1

A(t)
Proportion of nodes
In state i=2

= 0.01, = 0.005, = 0.0001, = 0.0001, = 0.3, r = 0.1, = 0.0001
Sample Runs with N = 1000

\[
\beta = 0.01, \delta_A = 0.005, \delta_D = 0.0001, \alpha_0 = \alpha = 0.0001, h = 0.3, r = 0.1, \lambda = 0.0001
\]
Example: WiFi Collision Resolution Protocol

- \( N \) nodes, state = retransmission stage \( k \)

- Time is discrete, \( I(N) = 1/N \); mean field limit is an ODE

- Occupancy measure is \( M(t) = [M_0(t),...,M_k(t)] \) with \( M_k(t) \) = proportion of nodes at stage \( k \)

Example: HTTP Metastability

- N flows between hosts and servers
- Flow $n$ is OFF or ON
- Time is discrete, occupancy measure = proportion of ON flows
- At every time step, every flow switches state with proba matrix that depends on the proportion of ON flows

- $I(N) = 1$; Mean field limit is an iterated map (discrete time)

[Baccelli et al.(2004)Baccelli, Lelarge, and McDonald]

- Other Examples where the mean field limit is in discrete time:
  - TCP flows with a buffer in [Tinnakornrisuphap and Makowski(2003)]
  - Reputation System in [Le Boudec et al.(2007)Le Boudec, McDonald, and Mundinger]
Example: Age of Gossip
Example: Age of Gossip

- Mobile node state = (c, t)
  - c = 1 ... 16 (position)
  - t ∈ R⁺ (age)

- Time is continuous, I(N) = 1

- Occupancy measure is
  \( F_c(z,t) = \) proportion of nodes that at location c and have age \( \leq z \)

Extension to a Resource

- Model can be complexified by adding a global resource $R(t)$

- Slow: $R(t)$ is expected to change state at the same rate $I(N)$ as one object
  
  -> call it an object of a special class

- Fast: $R(t)$ is change state at the aggregate rate $N I(N)$
  
  -> requires special extensions of the theory

[Bordenave et al. (2007)]

[Benaïm and Le Boudec (2008)]
Contents

- Mean Field Interaction Model
- The Mean Field Limit
- Convergence to Mean Field
- The Decoupling Assumption
- Optimization
The Mean Field Limit

Under very general conditions (given later) the occupancy measure converges, in some sense, to a deterministic process, $m(t)$, called the mean field limit.

$$M^N \left( \frac{t}{t(N)} \right) \rightarrow m(t)$$

[Graham and Méléard(1994)] consider the occupancy measure $L^N$ in path space.

$$M^N(t) \overset{\text{def}}{=} \frac{1}{N} \sum_n \delta x^N_n(t)$$

$$L^N \overset{\text{def}}{=} \frac{1}{N} \sum_n \delta X^N_n$$
Mean Field Limit
$N = +\infty$

Stochastic system
$N = 1000$
Propagation of Chaos is Equivalent to Convergence to a Deterministic Limit

**Definition**

Let $X^N = (X_1^N, ..., X_N^N)$ be an exchangeable sequence of processes in $\mathcal{P}(S)$ and $m \in \mathcal{P}(S)$ where $S$ is metric complete separable. $(X^N)_N$ is $m$-chaotic iff for every $k$:

\[\mathcal{L}(X_1^N, ..., X_k^N) \to m \otimes ... \otimes m\] as $N \to \infty$.

**Theorem ([Sznitman(1991)])**

$(X^N)_N$ is $m$-chaotic then the occupancy measure

$M^N \overset{\text{def}}{=} \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n^N}$ converges in probability (and in law) to $m$.

If the occupancy measure converges in law to $m$ then $(X^N)_N$ is $m$-chaotic.
(Propagation of Chaos)
If the initial condition \((X^N_n(0))_{n=1...N}\) is exchangeable and there is mean field convergence then the sequence \((X^N_n)_{n=1...N}\) indexed by \(N\) is \(m\)-chaotic

\[k\) objects are asymptotically independent with common law equal to the mean field limit, for any fixed \(k\)

\[\mathcal{L} \left( X_1 \left( \frac{t}{l(N)} \right), ..., X_k \left( \frac{t}{l(N)} \right) \right) \rightarrow m(t) \otimes ... \otimes m(t)\]

(Decoupling Assumption)
(also called Mean Field Approximation, or Fast Simulation)
The law of one object is asymptotically as if all other objects were drawn randomly with replacement from \(m(t)\)
Example: Propagation of Chaos

At any time $t$

$$P(X_n(t) \not\rightarrow A') \approx A \left( \frac{t}{N} \right)$$

$$P(X_m(t) \not\rightarrow D', X_n(t) \not\rightarrow A') \approx D \left( \frac{t}{N} \right) A \left( \frac{t}{N} \right)$$

where $(D, A, S)$ is solution of ODE

Thus for large $t$:

- $\text{Prob (node } n \text{ is dormant)} \approx 0.3$
- $\text{Prob (node } n \text{ is active)} \approx 0.6$
- $\text{Prob (node } n \text{ is susceptible)} \approx 0.1$
Example: Decoupling Assumption

Let $p^N_j(t|i)$ be the probability that a node that starts in state $i$ is in state $j$ at time $t$:

$$p^N_j(t|i) = \mathbb{P}(X^N_n(t) = j | X^N_n(0) = i)$$

The decoupling assumptions says that

$$p^N_j(t/N|i) \approx p_j(t|i)$$

where $p(t|i)$ is a continuous time, non-homogeneous process

$$\frac{d}{dt} \bar{p}(t|i) = \bar{p}(t|i)^T A (\bar{\mu}(t))$$

$$\frac{d}{dt} \bar{\mu}(t) = \bar{\mu}(t)^T A (\bar{\mu}(t)) = F (\bar{\mu}(t))$$

[Tembine et al.(2009)Tembine, Le Boudec, El-Azouzi, and Altman]
[Le Boudec et al.(2007)Le Boudec, McDonald, and Mundinger]
The Two Interpretations of the Mean Field Limit

$m(t)$ is the approximation for large $N$ of
1. the occupancy measure $M^N(t)$
2. the state probability for one object at time $t$
Contents

- Mean Field Interaction Model
- The Mean Field Limit
- Convergence to Mean Field
- The Decoupling Assumption
- Optimization
The General Case

- Convergence to the mean field limit is very often true

- A general method is known [Sznitman(1991)]:
  - Describe original system as a markov system; make it a martingale problem, using the generator
  - Show that the limiting problem is defined as a martingale problem with unique solution
  - Show that any limit point is solution of the limiting martingale problem
  - Find some compactness argument (with weak topology)

- Requires knowing [Ethier and Kurtz(2005)]
Kurtz’s Theorem

- Original System is in discrete time and $I(N) \to 0$; limit is in continuous time
- State space for one object is finite

[Kurtz(1970), Sandholm(2006)] Let

$$f^N(m) \overset{\text{def}}{=} \frac{1}{I(N)} \mathbb{E} \left( M^N(k+1) - m \mid M^N(k) = m \right)$$

$$A^N(m) \overset{\text{def}}{=} \frac{1}{I(N)} \mathbb{E} \left( \| M^N(k+1) - m \| \mid M^N(k) = m \right)$$

$$B^N(m) \overset{\text{def}}{=} \frac{1}{I(N)} \mathbb{E} \left( \| M^N(k+1) - m \| \mathbf{1}_{\{\| M^N(k+1) - m \| > \delta_N \}} \mid M^N(k) = m \right)$$

- $\lim_N \sup_m \| f^N(m) - f(m) \| = 0$ for some $f$,
- $\sup_N \sup_m A^N(m) < \infty$
- $\lim_N \sup_m \| B^N(m) \| = 0$ with $\lim_{N \to \infty} \delta_N = 0$

$M^N(0) \to m_0$ in probability

Then $\sup_{0 \leq t \leq T} \mathbb{P} \left( \| M^N(t) - m(t) \| \right) \to 0$ in probability.
Discrete Time, Finite State Space per Object

- Refinement + simplification, with a fast resource
  [Benaïm and Le Boudec (2008), Ioannidis and Marbach (2009)]
  - Let $W^N(k)$ be the number of objects that do a transition in time slot $k$. Note that $\mathbb{E} (W^N(k)) = NI(N)$, where $I(N) \overset{\text{def.}}{=} \text{intensity}$. Assume

  $$\mathbb{E} \left( W^N(k)^2 \right) \leq \beta(N) \quad \text{with} \quad \lim_{N \to \infty} I(N) \beta(N) = 0$$

  - $M^N(0) \to m_0$ in probability
  - regularity assumption on the drift (generator)

Then $\sup_{0 \leq t \leq T} \mathbb{P} \left( \| M^N(t) - m(t) \| \right) \to 0$ in probability.

- When limit is non continuous:
  [Benaim et al. (2006) Benaim, Hofbauer, and Sorin]
Discrete Time, Enumerable State Space per Object

- State space is enumerable with discrete topology, perhaps infinite; with a fast resource

[Bordenave et al. (2007) Bordenave, McDonald, and Proutiere]
  - Probability that objects $i$ and $j$ do a transition in one time slot is $o(1/N)$
  - $M^N(0) \rightarrow m(0)$ in probability for the weak topology
  - $(X^N_1(0), ..., X^N_N(0))$ is exchangeable at time 0
  - Regularity assumption on the drift (generator)
  - Then $M^N$ is $m$-chaotic.

- Essentially: same as previous plus exchangeability at time 0
Mean field limit is in discrete time

\[ \lim_{N} I(N) = 1 \]

- Object \( i \) draws next state at time \( k \) independent of others with transition matrix \( K^N(M^N) \)
- \( M^N(0) \to m_0 \) a.s. [in probability]
- regularity assumption on the drift (generator)

Then \( \sup_{0 \leq k \leq K} P (\| M^N(k) - m(k) \|) \to 0 \) a.s. [in probability]
Continuous Time

- « Kurtz’s theorem » also holds in continuous time (finite state space)
- Graham and Méléard: A generic result for \textbf{general} state space (in particular non enumerable).

\[ l(N) = 1/N, \text{ continuous time.} \]

- Object \( i \) has a free evolution plus pairwise interactions.
- \( X^N_n(0)_{n=1,...,N} \) is iid with common law \( m_0 \)
- Generator of pairwise meetings is uniformly bounded in total variation norm
  e.g. if \( G \cdot \varphi(x, x') = \int \varphi(y, y')f(y, y'|x, x')dydy' \) then
  \[ \int |f(y, y'|x, x')| dydy' \leq \Lambda, \text{ for all } x, x' \]

Then there is propagation of chaos with explicit bounds in total variation over finite time intervals. Mean field independence holds.
Age of Gossip

- Every taxi has a state
  - Position in area \( c = 0 \ldots 16 \)
  - Age of last received information

- [Graham and Méléard 1997] applies, i.e. mean field convergence occurs for iid initial conditions

- [Chaintreau et al. (2009)] shows more, i.e. weak convergence of initial conditions suffices

\[
\forall c \in C, \quad \frac{\partial F_c(z, t)}{\partial t} + \frac{\partial F_c(z, t)}{\partial z} = \sum_{c' \neq c} \rho_{c',c} F_{c'}(z, t) - \left( \sum_{c' \neq c} \rho_{c,c'} \right) F_c(z, t) + (u_c(t|d) - F_c(z, t)) \left( 2\eta_c F_c(z, t) + \mu_c \right) + (u_c(t|d) - F_c(z, t)) \sum_{c' \neq c} 2\beta_{c,c'} F_{c'}(z, t)
\]

\[
\forall c \in C, \quad \forall t \geq 0, \quad F_c(0, t) = 0
\]

\[
\forall c \in C, \quad \forall z \geq 0, \quad F_c(z, 0) = F_c^0(z).
\]
The Bounded Confidence Model


- Discrete time. State space $=[0, 1]$. 
  $X_n^N(k) \in [0, 1]$ rating of common subject held by peer $n$

- Two peers, say $i$ and $j$ are drawn uniformly at random. 
  If $|X_i^N(k) - X_j^N(k)| > \Delta$ no change; else

\[
X_i^N(k + 1) = wX_i^N(k) + (1 - w)X_j^N(k), \\
X_j^N(k + 1) = wX_j^N(k) + (1 - w)X_i^N(k),
\]
PDF of Mean Field Limit
Is There Convergence to Mean Field?

- Intuitively, yes

- Discretized version of the problem:
  - Make set of ratings discrete
  - Generic results apply: number of meetings is upper bounded by 2
  - There is convergence for any initial condition such that $M^N(0) \rightarrow m_0$

- This is what matlab does.

- However, there can be no similar result for the real version of the problem
  - There are some initial conditions such that $M^N(0) \rightarrow m_0$ while there is not convergence to the mean field
  - There is convergence to mean field if initial condition is iid from $m_0$
Contents

- Mean Field Interaction Model
- The Mean Field Limit
- Convergence to Mean Field
- The Decoupling Assumption
- Optimization
Decoupling Assumption

- Is true when mean field convergence holds, i.e. almost always
- It is often used in stationary regime
Example

In stationary regime:
- Prob (node $n$ is dormant) $\approx 0.3$
- Prob (node $n$ is active) $\approx 0.6$
- Prob (node $n$ is susceptible) $\approx 0.1$

- Nodes $m$ and $n$ are independent

We are in the good case: the diagram commutes

Law of $M^N(t)$ $\overset{t \to +1}{\longrightarrow} \omega^N$

$N \to +1$ $\rightarrow$ $\delta_{\mu(t)}$ $\overset{t \to +1}{\longrightarrow} \delta_{m^*}$

$N \to +1$ $\rightarrow$ $\omega^N$
Counter-Example

- The ODE does not converge to a unique attractor (limit cycle)

- Assumption H does not hold; does the decoupling assumption still hold?

Same as before
Except for one parameter value

\( h = 0.1 \) instead of 0.3
Decoupling Assumption Does Not Hold Here
In Stationary Regime

- In stationary regime, $m(t) = (D(t), A(t), S(t))$ follows the limit cycle
- Assume you are in stationary regime (simulation has run for a long time) and you observe that one node, say $n=1$, is in state ‘A’
- It is more likely that $m(t)$ is in region $R$
- Therefore, it is more likely that some other node, say $n=2$, is also in state ‘A’

This is synchronization
Numerical Example

Stationary point of ODE

Mean of Limit of $\omega^N = \text{pdf of one node in stationary regime}$

pdf of node 2 in stationary regime, given node 1 is A

pdf of node 2 in stationary regime, given node 1 is S

pdf of node 2 in stationary regime, given node 1 is D
Numerical Results ($h = 0.1$).

<table>
<thead>
<tr>
<th>prob of state</th>
<th>D</th>
<th>A</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>given D</td>
<td>0.261</td>
<td>0.559</td>
<td>0.181</td>
</tr>
<tr>
<td>given A</td>
<td>0.152</td>
<td>0.583</td>
<td>0.264</td>
</tr>
<tr>
<td>given S</td>
<td>0.099</td>
<td>0.533</td>
<td>0.368</td>
</tr>
<tr>
<td>unconditional</td>
<td>0.154</td>
<td>0.565</td>
<td>0.281</td>
</tr>
</tbody>
</table>

### Simplified Analysis 2
**Decoupling Assumption (Stationary Regime)**

Solve for $(D,A,S)$

- 1. $\delta_D D + 2\lambda D^2 + \beta A \frac{D}{h + D} = (\alpha_0 + rD)S$
- 2. $2\lambda D^2 + \beta A \frac{D}{h + D} + \alpha S = \delta_A A$
- 3. $\delta_D D + \delta_A A = (\alpha_0 + rD)S + \alpha S$

- Solve for $(D,A,S)$
- Has a unique solution

```plaintext
<table>
<thead>
<tr>
<th>case</th>
<th>prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$D\delta_D$</td>
</tr>
<tr>
<td>2</td>
<td>$D\Lambda^{ND-1}_{N-1}$</td>
</tr>
<tr>
<td>3</td>
<td>$A\beta_{k+D}$</td>
</tr>
<tr>
<td>4</td>
<td>$A\delta_A$</td>
</tr>
<tr>
<td>5</td>
<td>$S(\alpha_0 + rD)$</td>
</tr>
<tr>
<td>6</td>
<td>$S\alpha$</td>
</tr>
</tbody>
</table>
```
Where is the Catch?

- Fluid approximation and fast simulation result say that nodes $m$ and $n$ are asymptotically independent.

- But we saw that nodes may not be asymptotically independent.

... is there a contradiction?
The Diagram Does Not Commute

\[
\mathbb{P}(X_1^N(t/N) = i \text{ and } X_1^N(t/N) = j) \xrightarrow{t \to \infty} \pi_{i,j}^N
\]

\[
\begin{align*}
N \to \infty & \\
\mu_i(t)\mu_j(t) & \\
\frac{1}{T} \int_0^T \mu_i(t)\mu_j(t) \, dt & \rightarrow N \to \infty
\end{align*}
\]

- For large \( t \) and \( N \):

\[
\mathbb{P}(X_1^N(t/N) = i \text{ and } X_1^N(t/N) = j) \approx \frac{1}{T} \int_0^T \mu_i(t)\mu_j(t) \, dt
\]

\[
\neq \left( \frac{1}{T} \int_0^T \mu_i(t) \, dt \right) \left( \frac{1}{T} \int_0^T \mu_j(t) \, dt \right)
\]

where \( T \) is the period of the limit cycle.
Generic Result for Stationary Regime

- **Original system** (stochastic):
  - \( (X^N(t)) \) is Markov, finite, discrete time
  - Assume it is irreducible, thus has a unique stationary proba \( \nu^N \)
  - Let \( \omega^N \) be the corresponding stationary distribution for \( M^N(t) \), i.e.
    \[
    P\left( M^N(t) = (x_1, \ldots, x_I) \right) = \omega^N(x_1, \ldots, x_I) \text{ for } x_i \text{ of the form } k/n, \ k \text{ integer}
    \]

- **Theorem** [Benaim]

**Theorem 3** The support of any limit point of \( \omega^N \) is a compact set included in the Birkhoff center of \( \Phi \).

Birkhoff Center: closure of set of points s.t. \( m \in \omega(m) \)

Omega limit: \( \omega(m) = \text{set of limit points of orbit starting at } m \)
Here: Birkhoff center = limit cycle $\cup$ fixed point

The theorem says that the stochastic system for large $N$ is close to the Birkhoff center, i.e. the stationary regime of ODE is a good approximation of the stationary regime of stochastic system.
M^N(t) is a Markov chain on E={(a, b, c) ≥ 0, a + b + c =1, a, b, c multiples of 1/N}

A. M^N(t) is periodic, this is why there is a limit cycle for large N.

B. For large N, the stationary proba of M^N tends to be concentrated on the blue cycle.

C. For large N, the stationary proba of M^N tends to a Dirac.

D. M^N(t) is not ergodic, this is why there is a limit cycle for large N.
Decoupling Assumption in Stationary Regime
Holds under (H)

- For large $N$ the **decoupling assumption** holds at any fixed time $t$
- It holds in stationary regime under assumption (H)
  - (H) ODE has a unique global stable point to which all trajectories converge
- Otherwise the **decoupling assumption** may not hold in stationary regime

- It has nothing to do with the properties at finite $N$
  - In our example, for $h=0.3$ the decoupling assumption holds in stationary regime
  - For $h=0.1$ it does not

- Study the ODE!
Existence and Unicity of a Fixed Point are not Sufficient for Validity of Fixed Point Method

- Essential assumption is (H) $\mu(\tau)$ converges to a unique $m^*$
- It is not sufficient to find that there is a unique stationary point, i.e. a unique solution to $F(m^*)=0$
- Counter Example on figure
  - ($X^N(t)$) is irreducible and thus has a unique stationary probability $\eta^N$
  - There is a unique stationary point ($=\text{fixed point}$) (red cross)
    - $F(m^*)=0$ has a unique solution
    - but it is not a stable equilibrium
  - The fixed point method would say here
    - Prob (node $n$ is dormant) $\approx 0.1$
    - Nodes are independent
  - ... but in reality
    - We have seen that nodes are not independent, but are correlated and \textit{synchronized}
Example: 802.11 with Heterogeneous Nodes

- Two classes of nodes with heterogeneous parameters (retransmission probability)

- Fixed point equation has a unique solution

- There is a limit cycle

[Cho2010]
Contents

- Mean Field Interaction Model
- The Mean Field Limit
- Convergence to Mean Field
- The Decoupling Assumption
- Optimization
Decentralized Control

- Game Theoretic setting; $N$ players, each player has a class, each class has a policy; each player also has a state;
  - Set of states and classes is fixed and finite
  - Time is discrete; a number of players plays at any point in time.
  - Assume similar scaling assumptions as before.

- [Tembine et al. (2009) Tembine, Le Boudec, El-Azouzi, and Altman] For large $N$ the game converges to a single player game against a population;

**Theorem 3.6.2** (Infinite $N$). Optimal strategies (resp. equilibrium strategies) exist in the limiting regime when $N \to \infty$ under uniform convergence and continuity of $\mathbb{R}^N \to \mathbb{R}$. Moreover, if $\{U^N\}$ is a sequence of $\varepsilon_N$—optimal strategies (resp. $\varepsilon_N$—equilibrium strategies) in the finite regime with $\varepsilon_N \to \varepsilon$, then, any limit of subsequence $U^{\phi(N)} \to U$ is an $\varepsilon$—optimal strategies (resp. $\varepsilon$—equilibrium) for game with infinite $N$. 
Centralized Control

- [Gast et al. (2010)] Gast, Gaujal, and Le Boudec
- Markov decision process
  - Finite state space per object, discrete time, $N$ objects
  - Transition matrix depends on a control policy
  - For large $N$ the system without control converges to mean field

- Mean field limit
  - ODE driven by a control function

- Theorem: under similar assumptions as before, the optimal value function of MDP converges to the optimal value of the limiting system

- The result transforms MDP into fluid optimization, with very different complexity
Conclusion

- Mean field models are frequent in large scale systems

- Writing the mean field equations is simple and provides a first order approximation

- Mean field is much more than a fluid approximation: decoupling assumption / fast simulation

- Decoupling assumption in stationary regime is not necessarily true.

- Mean field equations may reveal emerging properties

- Control on mean field limit may give new insights
References


