Computational methods for nonlocal mean field games with applications

Levon Nurbekyan

UCLA

Joint work with S. Liu, M. Jacobs, W. Li, S. Osher, and J. Saúde

The problem

We are interested in developing computational methods for

$$\begin{cases} -\phi_t + H(t, x, \nabla\phi, \nabla^2\phi) = f\left(x, \rho(x,t), \int_\Omega K(x,y)\rho(y,t)dy\right) \\ \rho_t - \sum_i \partial_{x_i}(\rho \nabla_{p_i} H) + \sum_{ij} \partial_{x_i x_j}(\rho \partial_{M_{ij}} H) = 0 \\ \rho(x,0) = \rho_0(x),\ \phi(x,1) = g\left(x, \rho(x,1), \int_\Omega S(x,y)\rho(y,1)dy\right), \end{cases}$$

- The source term, and the boundary condition of HJB model the interactions between agents.
- The nonlocal interaction terms

$$\int_\Omega K(x,y)\rho(y,t)dy, \quad \int_\Omega S(x,y)\rho(y,1)dy$$

make the problem challenging from computational perspective. Indeed, non-singular K, S yield dense systems on a discrete level.

Existing numerical methods

There are number of general-purpose numerical methods that handle the system above.

- Newton's method [ACD10, Ach13, ACCD13]
- Semi-Lagrangian methods [CS12, CS14, CS15]
- ADMM (Brenier-Benamou) [BC15, BCS17] for *potential* MFG
- PDHG [BnAKS18, BnAKK$^+$19] for *potential* MFG
- HJB in density-space [CLOY19] for *potential* MFG
- Monotone flows [AFG17]

However, these methods yield dense systems on the discrete level when the interactions are nonlocal. Thus, the algorithms become computationally expensive and not amenable to parallelization techniques.

Goal

We aim at developing computational framework that

- yields sparse systems by encoding interactions in a small number of *coefficients*
- yields computational cost that is on par with algorithms for local couplings
- suits well the Lagrangian framework
- is compatible with existing convex optimization techniques and numerical methods when interactions are of mixed type
- extends to the non-potential setting
- provides modeling framework for nonlocal problems

The references for our method are [Nur18, NS18, LJL^{+}20].

The method of coefficients

For concreteness and to illustrate the ideas, we consider the following system

$$\begin{cases} -\phi_t + H(x, \nabla\phi) = \int_\Omega K(x,y)\rho(y,t)dy \\ \rho_t - \nabla \cdot (\rho \nabla_p H(x, \nabla\phi)) = 0 \\ \rho(x,0) = \rho_0(x), \ \phi(x,1) = g(x) \end{cases}$$

Our discussion will be formal. However, one can prove rigorous results under, for instance, the following assumptions: $\Omega = \mathbb{T}^d, H \in C^2$, and

$$\frac{1}{C} I_d \leq \nabla^2_{pp} H(x,p) \leq C I_d, \quad -C(1+|p|^2) \leq \nabla_x H(x,p) \cdot p$$

Furthermore, $\rho_0 \in L^\infty(\mathbb{T}^d) \cap \mathcal{P}(\mathbb{T}^d), \ g \in C^2(\mathbb{T}^d)$ and $K \in C^2(\mathbb{T}^d \times \mathbb{T}^d)$.

The method of coefficients

The key idea is to rewrite $\int_\Omega K(x,y)\rho(y,t)dy$ in a "Fourier" space. More precisely, suppose that

$$K(x,y)=\sum_{i,j=1}^{r} k_{ij} f_i(x) f_j(y),$$

where $\{f_i\}_{i=1}^r \subset C^2(\Omega)$ is some family of functions.

Remark. In general, K may not have this form. In such cases, we approximate K with kernels of such form.

Key observation. For any ρ we *a priori* have that

$$\int_\Omega K(x,y)\rho(y,t)dt=\sum_{i=1}^{r} a_i(t) f_i(x),$$

where

$$a_i(t)=\sum_{j=1}^{r} k_{ij}\int_\Omega f_j(y)\rho(y,t)dy.$$

The method of coefficients

$$\begin{cases} -\phi_t + H(x, \nabla\phi) = \int_\Omega K(x,y)\rho(y,t)dy \\ \rho_t - \nabla\cdot(\rho\nabla_p H(x,\nabla\phi)) = 0 \\ \rho(x,0) = \rho_0(x),\ \phi(x,1) = g(x) \end{cases}$$

Therefore, the HJB equation becomes

$$\begin{cases} -\phi_t + H(x, \nabla\phi) = \sum_{i=1}^r a_i(t) f_i(x) \\ \phi(x,1) = g(x), \end{cases}$$

and to solve the MFG, we need to find the *unknown coefficients* $\{a_i(t)\}_{i=1}^r$.
Of course, given $\{a_i(t)\}_{i=1}^r$, the measure ρ_a corresponding to these set of coefficients must satisfy the compatibility condition

$$a_i(t) = \sum_{j=1}^r k_{ij} \int_\Omega f_j(y)\rho_a(y,t)dy.$$

The method of coefficients

Summarizing, we search for *unknown coefficients* $\{a_i(t)\}_{i=1}^r$ such that

$$a_i(t) = \sum_{j=1}^r k_{ij} \int_\Omega f_j(y)\rho_a(y,t)dy,$$

where ρ_a is the distributional solution of

$$\begin{cases} \rho_t - \nabla \cdot (\rho \nabla_p H(x, \nabla\phi_a)) = 0 \\ \rho(x,0) = \rho_0(x), \end{cases}$$

and ϕ_a is the viscosity solution of

$$\begin{cases} -\phi_t + H(x, \nabla\phi) = \sum_{i=1}^r a_i(t) f_i(x) \\ \phi(x,1) = g(x). \end{cases}$$

Remarks

- The coefficients $\{a_i\}$ contain all the information about the interactions, and there is no need to assemble $\int_\Omega K(x, y)\rho(y, t)dy$: we just need to keep track of $\{a_i\}$.
- As we shall see below, $\{a_i\}$ are variational; that is,
  - when $\rho \mapsto \int_\Omega K(x, y)\rho(y)dy$ is monotone, these are *zeroes* of a monotone inclusion,
  - when $\rho \mapsto \int_\Omega K(x, y)\rho(y)dy$ is monotone and K is symmetric, these are solutions of a *convex optimization problem*.

The last two observations will be the basis for computational methods that we develop.

A derivative formula

Recall that we have to solve

$$a_i(t) = \sum_{j=1}^{r} k_{ij} \int_\Omega f_j(y)\rho_a(y,t)dy,$$

Key idea. Search for gradients! Since $\{a_i\}$ are our parameters, we may try to see how ϕ_a varies when $\{a_i\}$ vary.

Theorem

[NS18, Theorem 2.3] The functional $a \mapsto \int_\Omega \phi_a(x,0)\rho_0(x)dx$ is concave and everywhere Fréchet differentiable. Moreover,

$$\frac{\delta}{\delta a_i} \int_\Omega \phi_a(x,0)\rho_0(x)dx = \int_\Omega f_i(x)\rho_a(x,\cdot)dx,\ 1 \le i \le r.$$

Remarks

- This formula is not surprising from control and constrained optimization perspectives. Indeed, we aim at evaluating the gradient of $a \mapsto \int_\Omega \phi_a(x,0)\rho_0(x)dx$ where (ϕ_a, a) are coupled via
$$\begin{cases} -\phi_t + H(x,\nabla\phi) = \sum_{i=1}^r a_i(t) f_i(x) \\ \phi(x,1) = g(x). \end{cases}$$
The *adjoint method* routine is in considering the adjoint of the linearized equation with respect to ϕ,
$$\begin{cases} \rho_t - \nabla \cdot (\rho \nabla_p H(x,\nabla\phi_a)) = 0 \\ \rho(x,0) = \rho_0(x), \end{cases}$$
and evaluating the gradient using the solution ρ_a. This method is a workhorse for PDE control and inverse problems.

- When $a \mapsto \int_\Omega \phi_a(x,0)\rho_0(x)dx$ is a value function; that is,

$$\int_\Omega \phi_a(x,0)\rho_0(x)dx = \inf_\zeta J(a,\zeta),$$

  the adjoint method reduces to the *envelope formula*

$$\frac{\delta}{\delta a}\int_\Omega \phi_a(x,0)\rho_0(x)dx = \partial_a J(a,\zeta_a),$$

  where $\zeta_a \in \arg\min_\zeta J(a,\zeta)$. The formal calculation is as follows:

$$\frac{d}{da}J(a,\zeta_a) = \partial_a J(a,\zeta_a) + \partial_\zeta J(a,\zeta_a)\partial_a \zeta_a = \partial_a J(a,\zeta_a)$$

  because $\partial_\zeta J(a,\zeta_a) = 0$ from a first-order optimality condition for ζ_a.
- This previous perspective is very fruitful.

Derivation via envelope formula

We want to derive

$$\frac{\delta}{\delta a_i} \int_\Omega \phi_a(x,0)\rho_0(x)dx = \int_\Omega f_i(x)\rho_a(x,\cdot)dx, \;\; 1 \le i \le r.$$

Going from Lagrangian to Eulerian coordinates in the control problem, we obtain

$$\int_\Omega \phi_a(x,0)\rho_0(x)dx = \inf_{\rho,v} \int_0^1 \int_\Omega \left(L(x,v) + \sum_i a_i(t) f_i(x) \right) \rho(x,t)dxdt + \int_\Omega g(x)\rho(x,1)dx$$

$$\text{s.t. } \rho_t + \nabla \cdot (\rho v) = 0, \quad \rho(x,0) = \rho_0(x)$$

- From here, we immediately obtain that $a \mapsto \int_\Omega \phi_a(x,0)\rho_0(x)dx$ is concave.

Derivation via envelope formula

Furthermore, we have that

$$\int_\Omega \phi_a(x,0)\rho_0(x)dx = \int_0^1 \int_\Omega \left(L(x,v_a) + \sum_i a_i(t) f_i(x) \right) \rho_a(x,t) dxdt + \int_\Omega g(x)\rho_a(x,1)dx$$

where $v_a = -\nabla_p H(x, \nabla\phi_a)$, and $\partial_t \rho_a - \nabla \cdot (\rho_a \nabla_p H(x, \nabla\phi_a)) = 0$, $\rho_a(x,0) = \rho_0(x)$.
Therefore, we apply the envelope formula and differentiate w.r.t. a ignoring the dependence of (ρ_a, v_a) on a:

$$\frac{\delta}{\delta a_i} \int_\Omega \phi_a(x,0)\rho_0(x)dx = \int_\Omega f_i(x)\rho_a(x,\cdot)dx, \; 1 \le i \le r.$$

Going back to the MFG system

Denote by $\mathbf{K} = (k_{ij})_{i,j=1}^r$. We can assume that $\mathbf{K}$ is invertible.

Theorem

[NS18, Theorem 3.1]

i. (ϕ, ρ) *is a solution of MFG iff* $(\phi, \rho) = (\phi_a, \rho_a)$ *for* $a \in C([0,1]; \mathbb{R}^r)$ *such that* $a = \mathbf{K} \frac{\delta}{\delta a} \int_\Omega \phi_a(x, 0) \rho_0(x) dx$.

ii. *If* $\mathbf{K}$ *is monotone, then the problem is equivalent to finding a zero of a monotone operator* $a \mapsto \mathbf{K}^{-1} a - \frac{\delta}{\delta a} \int_\Omega \phi_a(x, 0) \rho_0(x) dx$.

iii. *If* $\mathbf{K}$ *is also symmetric, the problem is equivalent to the convex program*

$$\inf_{a \in C([0,1];\mathbb{R}^r)} \frac{\langle \mathbf{K}^{-1} a, a \rangle}{2} - \int_\Omega \phi_a(x, 0) \rho_0(x) dx,$$

where $\langle a, b \rangle = \sum_{i=1}^r \int_0^1 a_i(t) b_i(t) dt$ *for* $a, b \in C([0,1]; \mathbb{R}^r)$.

Remarks

- $\mathbf{K}$ is monotone iff $\rho \mapsto \int_\Omega K(x,y)\rho(y)dy$ is monotone. The latter is essential for uniqueness of solutions.
- $\mathbf{K}$ is symmetric iff K is symmetric, that is, $K(x,y) = K(y,x)$.
- In the monotone setting, we can apply powerful convex optimization techniques to find a.
- Formula

$$\frac{\delta}{\delta a_i} \int_\Omega \phi_a(x,0)\rho_0(x)dx = \int_\Omega f_i(x)\rho_a(x,\cdot)dx, \; 1 \leq i \leq r.$$

  is critical for update rules of a. Indeed,
  - start from some a and produce optimal controls v_a,
  - apply v_a and produce ρ_a
  - update a by $a_i(t) \leftarrow a_i(t) - h\sum_{j=1}^{r}(\mathbf{K}^{-1})_{ij}a_j(t) + h\int_\Omega f_i(x)\rho_a(x,t)dx, \; \forall i,$ where h is the step-size.

Remarks

- As usual, explicit descent steps can be replaced by implicit proximal steps.
- Given a, we do not need to solve perfectly for (v_a, ρ_a) to calculate descent directions. Instead, we can produce approximate solutions that improve as the optimization goes on. As we shall see, primal-dual methods achieve precisely this.
- We can work in both Eulerian and Lagrangian settings. In the latter case, we have that

$$\frac{\delta}{\delta a_i} \int_\Omega \phi_a(x,0)\rho_0(x)dx = \int_\Omega f_i(z_a(x,t))\rho_0(x,\cdot)dx, \ 1 \leq i \leq r,$$

  where z_a are the optimal trajectories corresponding to a; that is, $z_a(x,0) = x$, and $\dot{z}_a = v_a(z_a, t)$ [NS18]. Therefore, we can go after *high-dimensional problems*.

Remarks

- Applying our framework, virtually any HJB solver or a single-agent trajectory planning algorithm can be augmented to solve an MFG problem with nonlocal couplings. Indeed, given a, we just need to produce optimal controls v_a and *correct* a-s, and so on.
- The stability theory for nonlocal MFG yields that once we produce an approximation $K_r(x,y)=\sum_{i,j=1}^{r} k_{ij} f_i(x) f_j(y)$ for $K(x,y)$, we obtain an approximation of the original problem that is uniform across all discretizations.
- Therefore, for a fixed r, there is no computational burden due to nonlocal couplings when the mesh is fine.

Remarks

- In the potential case, our formulation is the Lasry-Lions optimal control formulation written in Fourier coordinates [LL07]:

$$\inf_{\alpha} \int_0^1 \mathcal{F}^*(\alpha(\cdot,t))dt - \int_\Omega \phi(x,0)\rho_0(x)dx$$
$$\text{s.t. } -\phi_t + H(x,\nabla\phi) = \alpha, \quad \phi(x,1) = g(x),$$

where $\mathcal{F}(\rho) = \frac{1}{2}\int_\Omega K(x,y)\rho(x)\rho(y)dxdy$.
When $\mathcal{F}(\rho) = \int_\Omega F(\rho)$ then $\mathcal{F}^*(\alpha) = \int_\Omega F^*(\alpha)$. For the nonlocal case though, there is no formula for $\mathcal{F}^*$ on the continuum level unless you pass to Fourier coordinates: $\alpha(x,t) = \sum_{i=1}^r a_i(t)f_i(x)$. In this case,

$$\mathcal{F}^*(\alpha) = \frac{\langle \mathbf{K}^{-1}a, a\rangle}{2}$$

A PDHG algorithm

The value function representation of $a \mapsto \int_\Omega \phi_a(x,0)\rho_0(x) = J(a)$ allows us to develop convex optimization methods to find a. Let us start with the potential case: $\mathbf{K} = \mathbf{K}^\top$. In this case, we have

$$
\begin{aligned}
&\inf_a \frac{\langle \mathbf{K}^{-1} a, a\rangle}{2} - \int_\Omega \phi_a(x,0)\rho_0(x)dx \\
=&\inf_a \sup_{\rho_t+\nabla\cdot(\rho v)=0} \frac{\langle \mathbf{K}^{-1} a, a\rangle}{2} - \int_0^1 \int_\Omega \left(L(x,v) + \sum_i a_i(t) f_i(x) \right) \rho dxdt \\
&+ \int_\Omega g(x)\rho(x,1)dx \\
=&\inf_a \sup_{\rho_t+\nabla\cdot m=0} \frac{\langle \mathbf{K}^{-1} a, a\rangle}{2} - \int_0^1 \int_\Omega \rho L\left(x, \frac{m}{\rho}\right) + \sum_i a_i(t) f_i(x)\rho dxdt \\
&+ \int_\Omega g(x)\rho(x,1)dx
\end{aligned}
$$

A PDHG algorithm

$$
\begin{aligned}
&= \inf_{\substack{\phi(x,1)=g \\ a}} \sup_{\rho,m} \left\{ \frac{\langle \mathbf{K}^{-1} a, a\rangle}{2} - \int_\Omega \phi(x,0)\rho_0(x)dx \right. \\
&\quad - \int_\Omega \int_0^1 (\rho\phi_t + m\cdot\nabla\phi)\, dxdt \\
&\quad \left. - \int_\Omega \int_0^1 \rho \left(L\left(x, \frac{m}{\rho}\right) + \sum_{i=1}^r a_i(t) f_i(x) \right) dxdt \right\} \\
&= \inf_{\substack{\phi(x,1)=g \\ a}} \sup_{\rho,m} \; \mathcal{L}(\phi, a, \rho, m)
\end{aligned}
$$

Note that $(\phi, a) \mapsto \mathcal{L}(\phi, a, \rho, m)$ is convex, $(\rho, m) \mapsto \mathcal{L}(\phi, a, \rho, m)$ is concave, and the coupling between (ϕ, a) and (ρ, m) is bilinear. Thus, we can apply PDHG [CP11, CP16] to solve this problem.

The algorithm

For step-sizes $\tau_{\nabla\phi}, \tau_{\phi_t}, \tau_{\phi(0)}, \tau_\rho, \tau_m > 0$, and current iterates $(a^k, \phi^k, \rho^k, m^k, \bar{a}^k, \bar{\phi}^k)$ the update rules for PDHG are

$$
\begin{cases}
(\rho^{k+1}, m^{k+1}) & = \underset{\rho,m}{\operatorname{argmax}}\ \mathcal{L}(\bar{\phi}^k, \bar{a}^k, \rho, m) - \frac{1}{2\tau_\rho}\|\rho - \rho^k\|^2_{L^2_{x,t}} \\
& \quad -\frac{1}{2\tau_m}\|m - m^k\|^2_{L^2_{x,t}} \\
(a^{k+1}, \phi^{k+1}) & = \underset{a,\phi}{\operatorname{argmin}}\ \mathcal{L}(\phi, a, \rho^{k+1}, m^{k+1}) \\
& \quad +\frac{1}{2\tau_{\phi(0)}}\|\phi(\cdot,0) - \phi^k(\cdot,0)\|^2_{L^2_x} + \frac{1}{2\tau_{\nabla\phi}}\|\nabla\phi - \nabla\phi^k\|^2_{L^2_{x,t}} \\
& \quad +\frac{1}{2\tau_{\phi_t}}\|\phi_t - \phi_t^k\|^2_{L^2_{x,t}} + \frac{1}{2\tau_a}\|a - a^k\|^2_{L^2_t} \\
(\bar{a}^{k+1}, \bar{\phi}^{k+1}) & = 2(a^{k+1}, \phi^{k+1}) - (a^k, \phi^k)
\end{cases}
$$

Correct choice of spaces

- As illustrated in [JLLO19, JL19], the choices of spaces for variables are crucial when applying PDHG. Correct choices render algorithms with grid-size-independent convergence rates.
- The norm of the bilinear coupling

$$\left|\int_\Omega \int_0^1 (\rho\phi_t + m \cdot \nabla\phi)\, dxdt\right| \leq \|(\rho, m)\|_{L^2} \cdot \|\phi\|_{H^1}$$

  is finite if we choose L^2 norm for (ρ, m) and H^1 norm for ϕ.
- Therefore, to obtain a grid-independent convergence rates, we must choose H^1 norm for ϕ.

The updates for (ρ, m)

First-order optimality conditions for (ρ, m) yield

$$\begin{cases} \nabla_v L\left(x, \frac{m}{\rho}\right) \cdot \frac{m}{\rho} - L\left(x, \frac{m}{\rho}\right) - \frac{\rho-\rho^k}{\tau_\rho} = \bar{\phi}_t^k + \sum_{i=1}^r \bar{a}_i^k(t) f_i(x) \\ \nabla_v L\left(x, \frac{m}{\rho}\right) + \frac{m-m^k}{\tau_m} = \nabla \bar{\phi}^k \end{cases}$$

- Note that we obtain a decoupled one-dimensional convex optimization problems at the grid-points. Therefore the proximal update for (ρ, m) can de performed efficiently in parallel.
- Direct applications of existing methods to nonlocal problems do not possess this property.
- The price that we pay are explicit updates for a small number of coefficients $\{a_i\}$.

The updates for (a, ϕ)

The first-order optimality conditions for (a, ϕ) yield

$$a^{k+1} = (\tau_a \mathbf{K}^{-1} + \mathbf{I})^{-1} \left(a^k + \tau_a \left(\int_\Omega f_i(x) \rho^{k+1}(x,t) dx \right)_{i=1}^{r} \right),$$

and a space-time elliptic equation for ϕ,

$$\begin{cases} \frac{\phi_{tt}}{\tau_{\phi_t}} + \frac{\Delta \phi}{\tau_{\nabla \phi}} = \rho_t^{k+1} + \nabla \cdot m^{k+1} + \frac{\phi_{tt}^k}{\tau_{\phi_t}} + \frac{\Delta \phi^k}{\tau_{\nabla \phi}} \quad \text{in} \quad \Omega \times (0,1) \\ \frac{\phi_t(x,0)}{\tau_{\phi_t}} - \frac{\phi(x,0)}{\tau_{\phi(0)}} = \rho^{k+1}(x,0) - \rho_0(x) + \frac{\phi_t^k(x,0)}{\tau_{\phi_t}} - \frac{\phi^k(x,0)}{\tau_{\phi(0)}} \quad \text{in} \quad \Omega \\ \phi(x,1) = g(x) \quad \text{in} \quad \Omega \\ \frac{\partial \phi(x,t)}{\partial \nu} = \frac{\partial \phi^k(x,t)}{\partial \nu} + \tau_{\nabla \phi} m^{k+1} \cdot \nu \quad \text{in} \quad \partial\Omega \times (0,1), \end{cases}$$

that can be efficiently solved via Fast Fourier Transform (FFT).

Kernels and bases

- Our method is flexible in the choice of $\{f_i\}$: trigonometric functions, polynomials, etc.
- Kernel methods in machine-learning provide a very suitable framework for choosing K and $\{f_i\}$
- Recall that a generic agent solves

$$\inf_u \int_t^1 \left\{ L(z(s), u(s)) + \int_\Omega K(z(s), y)\rho(y, t)dy \right\} ds + g(z(T))$$

  Therefore, in a monotone regime, $K(x, y)$ is a *similarity measure* between positions x and y that agents try to minimize.
- In other words, agents avoid locations that have similar features.

Kernels and bases

- Kernel methods in ML utilize precisely these types of kernels.
- The simplest example of K is the inner product, $K(x,y) = x \cdot y$, which is amenable to rigorous mathematical analysis. Natural extensions of the inner product are *positive definite symmetric (PDS)* kernels.
- $K : (x,y) \mapsto K(x,y)$ is a PDS kernel if $(K(x^i,x^j))_{i,j=1}^m$ is symmetric positive semidefinite matrix for all $\{x^i\}_{i=1}^m \subset \mathbb{R}^d$.
- Assume K is a continuous PDS. Then for arbitrary $\rho_k = \frac{1}{N}\sum_i w_k^i \delta_{x^i},\ k=1,2$ we have that

$$\int_{\Omega^2} K(x,y) d(\rho_2(x)-\rho_1(x)) d(\rho_2(y)-\rho_1(y))$$
$$= \sum_{i,j} K(x^i,x^j)(w_2^i - w_1^i)(w_2^j - w_1^j) \geq 0,$$

and hence $\rho \mapsto \int_\Omega K(x,y)\rho(y)dy$ is monotone.

Kernels and bases

- A remarkable fact about PDS kernels is that they are inner products in a suitably chosen Hilbert space

$$K(x,y)=\langle f(x), f(y)\rangle_{\mathcal{H}}, \quad \forall x,y$$

- $f(x)$ is called the *feature vector* of x. If $\mathcal{H}$ is separable, we have that $f(x)=(f_1(x), f_2(x), \cdots, f_n(x), \cdots)$, and

$$\begin{aligned} K(x,y)=&\langle f(x), f(y)\rangle_{\mathcal{H}}=\langle \sum_i f_i(x)e_i, \sum_i f_i(y)e_i\rangle_{\mathcal{H}} \\ =&\sum_{i,j}\langle e_i, e_j\rangle_{\mathcal{H}} f_i(x)f_j(y)=\sum_{i,j} k_{ij} f_i(x)f_j(y). \end{aligned}$$

Therefore, we obtain the representation we need!

Examples

- **Maximal spread.** $K(x,y) = 2\sum_{i=1}^d \lambda_i x_i y_i = \sum_{i=1}^d f_i(x) f_i(y)$, where $f_i(x) = \sqrt{2\lambda_i} x_i$. In this case we obtain $\mathbf{K} = \mathbf{I}$, and the a-update is trivial

$$a_i^{k+1}(t) = \frac{\tau_a \int_\Omega f_i(x) \rho^{k+1}(x,t) dx + a_i^k(t)}{\tau_a + 1}$$

- **Gaussian repulsion.** $K(x,y) = \mu \prod_{i=1}^d \exp\left(-\frac{|x_i - y_i|^2}{2\sigma_i^2}\right)$. In this case, we have that

$$K(x,y) = \sum_{\alpha_1,\alpha_2,\cdots,\alpha_d \geq 0} f_{\alpha_1,\alpha_2,\cdots,\alpha_d}(x) f_{\alpha_1,\alpha_2,\cdots,\alpha_d}(y),$$

where

$$f_{\alpha_1,\alpha_2,\cdots,\alpha_d}(x) = \sqrt{\mu} e^{-\sum_{i=1}^d \frac{|x_i|^2}{2\sigma_i^2}} \prod_{i=1}^d \frac{x_i^{\alpha_i}}{\sigma_i^{\alpha_i} \alpha_i!}, \quad \alpha_1, \alpha_2, \cdots, \alpha_d \geq 0.$$

Again, we obtain $\mathbf{K} = \mathbf{I}$.

Examples

- **Differential operators.** In [ACD10] the authors consider $V[\rho] = \mu(I-\Delta)^{-2}\rho$ on $\mathbb{T}^d$. One has that

$$V[\rho] = \int_{\mathbb{T}^d} \Gamma(x-y)\rho(y)dy,$$

where $(I-\Delta)^2\Gamma = \mu\delta_0$. So $K(x,y) = \Gamma(x-y)$, and a suitable choice for $\{f_i\}$ are the trigonometric functions that diagonalize K; that is,

$$K(x,y) = \sum_{\alpha\geq 0} f_\alpha^{\cos}(x) f_\alpha^{\cos}(y) + \sum_{\alpha>0} f_\alpha^{\sin}(x) f_\alpha^{\sin}(y),$$

where $f_\alpha^{\cos}(x) = \sqrt{\gamma_\alpha}\cos(2\pi\alpha\cdot x)$, $f_\alpha^{\sin}(x) = \sqrt{\gamma_\alpha}\sin(2\pi\alpha\cdot x)$, and $\Gamma(x) = \sum_{\alpha\geq 0}\gamma_\alpha\cos(2\pi\alpha\cdot x)$. Again, we obtain $\mathbf{K} = \mathbf{I}$.

Non-potential case

- When K is non-symmetric we obtain a non-potential MFG. In this case, we need to solve the monotone inclusion

$$0 \in Ta - \partial_a J(a)$$

where $Ta = \mathbf{K}^{-1}a$, and $J(a) = \int_\Omega \phi_a(x,0)\rho_0(x)dx$. Nevertheless, PDHG algorithms extend to monotone inclusions [CP12].

- We just need to perform proximal steps in a as follows

$$a^{k+1} = (\tau_a T + \mathbf{I})^{-1}\left(a^k + \tau_a \left(\int_\Omega f_i(x)\rho^{k+1}(x,t)dx\right)_{i=1}^r\right)$$

Mixed couplings

- Assume that we want to solve

$$\begin{cases} -\phi_t + H(x, \nabla\phi) = f(\rho(x,t)) + \int_\Omega K(x,y)\rho(y,t)dy \\ \rho_t - \nabla \cdot (\rho \nabla_p H(x, \nabla\phi)) = 0 \\ \rho(x,0) = \rho_0(x), \ \phi(x,1) = g(x), \end{cases}$$

  where f is some monotone function.

- Then, we introduce parameters $\{a_i(t)\}$ to handle the nonlocal case, and $\alpha(x,t)$ for the local part

$$\begin{cases} -\phi_t + H(x, \nabla\phi) = \alpha(x,t) + \sum_{i=1}^r a_i(t) f_i(x) \\ \phi(x,1) = g(x), \end{cases}$$

Mixed couplings

- Again, we denote by $\phi_{\alpha,a}$ the solution of

$$\begin{cases} -\phi_t + H(x, \nabla\phi) = \alpha(x,t) + \sum_{i=1}^{r} a_i(t) f_i(x) \\ \phi(x,1) = g(x), \end{cases}$$

and by $\rho_{\alpha,a}$ the solution of

$$\begin{cases} \rho_t - \nabla \cdot (\rho \nabla_p H(x, \nabla\phi_{\alpha,a})) = 0 \\ \rho(x,0) = \rho_0(x), \end{cases}$$

- The derivative formulas for $J_{\alpha,a} = \int_\Omega \phi_{\alpha,a}(x,0)\rho_0(x)dx$ in this case become

$$\partial_\alpha J = \rho_{\alpha,a}, \quad \partial_{a_i} J = \int_\Omega f_i(x) \rho_{\alpha,a}(x,\cdot)dx$$

Mixed couplings

- Then MFG can be written as

$$\begin{cases} a = \mathbf{K}\partial_a J \\ \alpha = f(\partial_\alpha J) \end{cases} \Leftrightarrow \begin{cases} \mathbf{K}^{-1}a - \partial_a J = 0 \\ (F^*)'(\alpha) - \partial_\alpha J = 0, \end{cases}$$

  where F^* is the convex dual of $F'(\rho) = f(\rho)$. Therefore, we again obtain monotone inclusions and can apply primal-dual algorithms.
- This means that we can handle convex point-wise constraints on ρ mixed with nonlocal couplings.
- For instance, $0 \leq \rho \leq 1$ corresponds to

$$F(\rho) = \mathbf{1}_{0 \leq \rho \leq 1}, \quad F^*(\alpha) = \max\{\alpha, 0\}.$$

  Proximal update steps for α are explicit in this case.

Extensions

- Stochastic problems
- General dynamics
- Nonlinear couplings $f\left(x, \rho(x,t), \int_\Omega K(x,y)\rho(y,t)dy\right)$

Numerical experiments

For numerical experiments see

- [LJL+20]
- [NS18]
- [Nur18]

Thank you for your attention!

Yves Achdou, Fabio Camilli, and Italo Capuzzo-Dolcetta.
Mean field games: convergence of a finite difference method.
SIAM J. Numer. Anal., 51(5):2585–2612, 2013.

Yves Achdou and Italo Capuzzo-Dolcetta.
Mean field games: numerical methods.
SIAM J. Numer. Anal., 48(3):1136–1162, 2010.

Yves Achdou.
Finite difference methods for mean field games.
In *Hamilton-Jacobi equations: approximations, numerical analysis and applications*, volume 2074 of *Lecture Notes in Math.*, pages 1–47. Springer, Heidelberg, 2013.

Noha Almulla, Rita Ferreira, and Diogo Gomes.
Two numerical approaches to stationary mean-field games.
Dyn. Games Appl., 7(4):657–682, 2017.

J.-D. Benamou and G. Carlier.

Augmented Lagrangian methods for transport optimization, mean field games and degenerate elliptic equations.
J. Optim. Theory Appl., 167(1):1–26, 2015.

J.-D. Benamou, G. Carlier, and F. Santambrogio.
Variational mean field games.
In *Active particles. Vol. 1. Advances in theory, models, and applications*, Model. Simul. Sci. Eng. Technol., pages 141–171. Birkhäuser/Springer, Cham, 2017.

L. Briceño Arias, D. Kalise, Z. Kobeissi, M. Laurière, Á. Mateos González, and F. J. Silva.
On the implementation of a primal-dual algorithm for second order time-dependent mean field games with local couplings.
In *CEMRACS 2017—numerical methods for stochastic models: control, uncertainty quantification, mean-field*, volume 65 of *ESAIM Proc. Surveys*, pages 330–348. EDP Sci., Les Ulis, 2019.

L. M. Briceño Arias, D. Kalise, and F. J. Silva.

Proximal methods for stationary mean field games with local couplings.
SIAM J. Control Optim., 56(2):801–836, 2018.

Yat Tin Chow, Wuchen Li, Stanley Osher, and Wotao Yin.
Algorithm for Hamilton–Jacobi equations in density space via a generalized Hopf formula.
J Sci Comput, 80(2):1195–1239, Aug 2019.

Antonin Chambolle and Thomas Pock.
A first-order primal-dual algorithm for convex problems with applications to imaging.
J. Math. Imaging Vision, 40(1):120–145, 2011.

Patrick L. Combettes and Jean-Christophe Pesquet.
Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators.
Set-Valued Var. Anal., 20(2):307–330, 2012.

Antonin Chambolle and Thomas Pock.

On the ergodic convergence rates of a first-order primal-dual algorithm.
Math. Program., 159(1-2, Ser. A):253–287, 2016.

Fabio Camilli and Francisco Silva.
A semi-discrete approximation for a first order mean field game problem.
Netw. Heterog. Media, 7(2):263–277, 2012.

E. Carlini and F. J. Silva.
A fully discrete semi-Lagrangian scheme for a first order mean field game problem.
SIAM J. Numer. Anal., 52(1):45–67, 2014.

Elisabetta Carlini and Francisco J. Silva.
A semi-Lagrangian scheme for a degenerate second order mean field game system.
Discrete Contin. Dyn. Syst., 35(9):4269–4292, 2015.

M. Jacobs and F. Léger.

A fast approach to optimal transport: the back-and-forth method.
Preprint, 2019.
arXiv:1905.12154 [math.OC].

Matt Jacobs, Flavien Léger, Wuchen Li, and Stanley Osher.
Solving large-scale optimization problems with a convergence rate independent of grid size.
SIAM J. Numer. Anal., 57(3):1100–1123, 2019.

Siting Liu, Matthew Jacobs, Wuchen Li, Levon Nurbekyan, and Stanley J. Osher.
Computational methods for nonlocal mean field games with applications, 2020.

Jean-Michel Lasry and Pierre-Louis Lions.
Mean field games.
Jpn. J. Math., 2(1):229–260, 2007.

Levon Nurbekyan and J. Saúde.

Fourier approximation methods for first-order nonlocal mean-field games.
Port. Math., 75(3-4):367–396, 2018.

Levon Nurbekyan.
One-dimensional, non-local, first-order stationary mean-field games with congestion: a Fourier approach.
Discrete Contin. Dyn. Syst. Ser. S, 11(5):963–990, 2018.