Splitting algorithms for non-smooth convex optimization[1]: Review and application to Mean Field Games[2]

L. M. Briceño-Arias
Universidad Técnica Federico Santa María

[1] with J. Deride, C. Vega, S. López
[2] with D. Kalise, F.J. Silva

Mean Field Games (MFG)[3]: Static example

- N players choose their positions on a set Q (compact).
- $\mathcal{P}(Q)$ is the set of Borel probability measures.
- They minimize their distance to a place $P \in Q$.
- Players are congestion-averse.
- The cost of player i can be modeled by

$$f_i(x_1, \ldots, x_i, \ldots, x_N) = \alpha|x_i - P| - \frac{\beta}{N-1}\sum_{j\neq i}|x_j - x_i|$$

$$= \alpha|x_i - P| - \beta\int_Q |x - x_i| \mathrm{d}\left(\frac{1}{N-1}\sum_{j\neq i}\delta_{x_j}\right)$$

$$= f\left(x_i, \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j}\right).$$

[3] J.-M. Lasry, P.-L. Lions. Mean field games. *Jpn. J. Math.* 2007
M. Huang, R. P. Malhamé, P. E. Caines. Large population stochastic dynamic games. *Commun. Inf. Syst.* 2006.

- Suppose that for each N, $(\bar{x}_1^N, \dots, \bar{x}_N^N)$ is a Nash equilibrium of the previous game, i.e., for every $i \in \{1, \dots, N\}$,

$$(\forall x_i \in Q) \quad f\left(\bar{x}_i^N, \frac{1}{N-1}\sum_{j\neq i}\delta_{\bar{x}_j^N}\right) \leq f\left(x_i, \frac{1}{N-1}\sum_{j\neq i}\delta_{\bar{x}_j^N}\right)$$

- Then, $\exists\ \bar{m} \in \mathcal{P}(Q)$ such that, up to some sub-sequence,

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{\bar{x}_i^N} \overset{*}{\rightharpoonup} \bar{m}.$$

- The equilibrium $\bar{m}$ satisfies the fixed point equation

$$\operatorname{supp}(\bar{m}) \subseteq \operatorname{argmin}\left\{f(x,\bar{m}) \mid x \in Q\right\}.$$

Dynamic & deterministic case

- Differential game with N players, where Player i minimizes

$$\int_0^T \left[\frac{|\alpha(t)|^2}{2} + f\Big(x_i(t), \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j(t)}\Big)\right] dt + g\left(x_i(T), \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j(T)}\right)$$

$$\text{s.t.} \quad \dot{x}_i(t) = \alpha(t) \quad \forall t \in [0,T],$$

$$x_i(0) = \bar{x}_{0,i}^N.$$

- Suppose that $(\bar{x}_1^N, \ldots, \bar{x}_N^N)$ is a Nash equilibrium and that

$$\frac{1}{N}\sum_{i=1}^N \delta_{\bar{x}_{0,i}^N} \overset{*}{\rightharpoonup} m_0.$$

Then for each $t \in [0,T]$, $\exists\, m(t) \in \mathcal{P}(Q)$ such that

$$\frac{1}{N}\sum_{i=1}^N \delta_{\bar{x}_i^N(t)} \overset{*}{\rightharpoonup} m(t)$$

Dynamic & deterministic case

- Any equilibrium m solves the MFG system

$$
\begin{aligned}
(HJB) \qquad -\partial_t u + \frac{|\nabla u|^2}{2} &= f(x, m(t)) \\
u(x,T) &= g(x, m(T)), \\
(FP) \quad \partial_t m - \operatorname{div}(m \nabla u) &= 0, \\
m(0) &= m_0.
\end{aligned}
$$

- At (x,t) the solution u of the HJB equation is given by

$$
\begin{aligned}
u(x,t) = \inf_{\alpha} \int_t^T \left[\frac{|\alpha(s)|^2}{2} + f(x(s), m(s)) \right] ds + g(x(T), m(T)) \\
\text{s.t.} \qquad \dot{x}(s) = \alpha(s) \qquad \forall s \in (t,T), \\
x(t) = x.
\end{aligned}
$$

Dynamic & stochastic case

$$\int_0^T \left[\frac{|\alpha(t)|^2}{2} + f\Big(x_i(t), \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j(t)}\Big)\right] dt + g\left(x_i(T), \frac{1}{N-1}\sum_{j\neq i}\delta_{x_j(T)}\right)$$

$$\text{s.t.} \qquad dx_i(t) = \alpha(t)dt + \sigma dW_i(t) \qquad \forall t\in[0,T],$$

$$x_i(0) = \bar{x}_{0,i}^N.$$

As before, $\frac{1}{N}\sum_{i=1}^N \delta_{\bar{x}_i^N(t)} \to m(t)$ which now solves ($t\in[0,T]$, $x\in Q$):

MFG

$$-\partial_t u - \frac{\sigma^2}{2}\Delta u + \frac{|\nabla u|^2}{2} = f(x, m(t))$$

$$u(x,T) = g(x, m(T))$$

$$\partial_t m - \frac{\sigma^2}{2}\Delta m - \operatorname{div}(m\nabla u) = 0$$

$$m(0) = m_0$$

Stationary MFG

- The ergodic counterpart of MFG is $(x \in Q)$

Stationary MFG (SMFG)

$$-\frac{\sigma^2}{2}\Delta u + \frac{|\nabla u|^2}{2} + \lambda = f(x,m),$$
$$-\frac{\sigma^2}{2}\Delta m - \operatorname{div}(m\nabla u) = 0,$$
$$\int_Q u(x)dx = 0, \quad m \geq 0, \ \int_Q m(x)dx = 1.$$

- The solution (u, m, λ) of SMFG describes the long time average of solutions (u^T, m^T) of MFG as $T \to \infty$ [4].

[4] P. Cardaliaguet, J.-M. Lasry, P.-L. Lions, A. Porretta, *Long time average of mean field games with a nonlocal coupling, SIAM J. Control Optim.*, 2013

Stationary MFG with local couplings

- In this talk we focus on local interactions $f(x, m) = f(x, m(x))$.
- We set $Q = \mathbb{T}^2$ and $\nu = \sigma^2/2$.
- We study numerical methods for solving

SMFG

$$
\begin{aligned}
-\nu\Delta u(x) + \frac{1}{2}|\nabla u(x)|^2 + \lambda = f(x, m(x)) \quad & \text{in } \mathbb{T}^2 \\
-\nu\Delta m(x) - \operatorname{div}\big(m(x)\nabla u(x)\big) = 0 \quad & \text{in } \mathbb{T}^2 \\
m \geq 0, \quad \int_{\mathbb{T}^2} m(x)dx = 1, \quad \int_{\mathbb{T}^2} u(x)dx = 0. &
\end{aligned}
$$

Discrete SMFG

DSMFG (Achdou & Capuzzo Dolcetta, 2010)

$$-\nu(\Delta_h u^h)_{i,j} + \frac{1}{2}|\widehat{[D_h u^h]}_{i,j}|^2 + \lambda^h = f(x_{i,j}, m^h_{i,j}) \quad \forall\, 0\le i,j\le N_h-1$$

$$-\nu(\Delta_h m^h)_{i,j} - \left(\mathrm{div}_h(m^h \widehat{[D_h u^h]})\right)_{i,j} = 0 \quad \forall\, 0\le i,j\le N_h-1$$

$$m^h_{i,j} \geq 0,\ h^2 \sum_{i,j} m^h_{i,j} = 1,\ \sum_{i,j} u^h_{i,j} = 0.$$

- $h > 0$, $N_h = 1/h$, $\mathcal{M}_h = \mathbb{R}^{N_h \times N_h}$, $\mathcal{W}_h = \mathbb{R}^{4(N_h \times N_h)}$.
- $\widehat{[D_h u]}_{i,j} = \left((D_1 u)^-_{i,j}, -(D_1 u)^+_{i-1,j}, (D_2 u)^-_{i,j}, -(D_2 u)^+_{i,j-1}\right) \in \mathbb{R}^4$, where $(D_1 u)_{i,j} := \frac{u_{i+1,j} - u_{i,j}}{h}$, $(D_2 u)_{i,j} := \frac{u_{i,j+1} - u_{i,j}}{h}$.
- Δ_h and div_h are linear operators defined by

$$(\Delta_h m)_{i,j} := -\frac{1}{h^2}(4m_{i,j} - m_{i+1,j} - m_{i-1,j} - m_{i,j+1} - m_{i,j-1})$$

$$\left(\mathrm{div}_h(w)\right)_{i,j} := (D_1 w^1)_{i-1,j} + (D_1 w^2)_{i,j} + (D_2 w^3)_{i,j-1} + (D_2 w^4)_{i,j}.$$

SMFG discretization

- If $\nu > 0$, $f(x, \cdot)$ is increasing and we suppose that the stationary system admits a unique classical solution, in Achdou, Camilli & Capuzzo Dolcetta (2013) the convergence of DSMFG (unif-L^2) to the unique solution to the stationary system as $h \to 0$ is proved.
- To solve the discretized system, Newton's method can be used (Achdou & Capuzzo Dolcetta, 2010; Achdou & Perez, 2012; Cacace & Camilli, 2016) if the initial guess is close enough to the solution.
- The performance of Newton's method depends heavily on the values of ν: for small values of ν the convergence is much slower and cannot be guaranteed in general since m^h can become negative.

Variational Approach

$$b(m,w)=\begin{cases}\frac{|w|^2}{2\,m}, & \text{if } m>0;\\ 0, & \text{if } (m,w)=(0,0);\\ +\infty, & \text{otherwise,}\end{cases} \qquad F(x,m)=\begin{cases}\int_0^m f(x,m')dm', & \text{if } m\geq 0;\\ +\infty, & \text{otherwise.}\end{cases}$$

The SMFG is (formally) the FOC of the optimization problem

Optimization Problem (P)

$$\inf_{m,w}\int_{\mathbb{T}^2}\left[b(m(x),w(x))+F(x,m(x))\right]dx$$

$$\text{s.t}\quad \begin{cases}-\nu\Delta m(x)+\operatorname{div}\big(w(x)\big)=0, & \text{in } \mathbb{T}^2\\ \int_{\mathbb{T}^2} m(x)dx=1,\end{cases}$$

where u and λ are Lagrange multipliers and $w=-m\nabla u$ (see Lasry & Lions, 2007).

Goal of this talk...

- Provide variational formulation (P_h) of DSMFG (Achdou & Capuzzo Dolcetta, 2010) in order to propose numerical approximations of the SMFG.
- Review of proximal splitting methods for solving (P_h).
  - Benamou & Carlier (2015) and Benamou, Carlier & Santambrogio (2017): FE discretization of the dynamic MFG via Augmented Lagrangian method (ADMM) case $\nu = 0$.
  - Andreev (2017): ADMM with preconditioners for $\nu > 0$ in dynamic case. Another discretization.
  - Papadakis, Peyré & Oudet (2014): optimal transport ($F = 0$) and $\nu = 0$ with centered grid.
- Connect first-order methods and fixed point iterations.
- Propose a new **projected** proximal splitting method for solving (P_h).
- Numerical experiments.

1. Motivation
2. (P_h)
3. Splitting algorithms: review
4. Split/unsplit approaches
5. Numerical experiences

Optimization Problem (P_h)

Discrete optimization problem (P_h)

$$\inf_{(m,w)\in\mathcal{M}_h\times\mathcal{W}_h} h^2 \sum_{i,j=0}^{N_h-1} \left[\hat{b}(m_{i,j}, w_{i,j}) + F(x_{i,j}, m_{i,j})\right]$$

$$\text{s.t.} \quad \begin{cases} -\nu(\Delta_h m)_{i,j} + (\mathrm{div}_h w)_{i,j} = 0, & \forall\, 0 \le i,j, \le N_h - 1 \\ h^2 \sum_{i,j} m_{i,j} = 1. \end{cases}$$

- $-\nu\Delta_h : \mathcal{M}_h \to \mathcal{M}_h$ and $\mathrm{div}_h : \mathcal{W}_h \to \mathcal{M}_h$ are linear.
- $\hat{b} \colon \mathbb{R}\times\mathbb{R}^4$ is given by

$$\hat{b} \colon (m,w) \mapsto \begin{cases} \dfrac{|w|^2}{2\,m}, & \text{if } m > 0,\ w \in K, \\ 0, & \text{if } (m,w) = (0,0), \\ +\infty, & \text{otherwise.} \end{cases}$$

- $K := \mathbb{R}_+ \times \mathbb{R}_- \times \mathbb{R}_+ \times \mathbb{R}_-$.

Existence of solutions to (P_h)

Qualification condition

There exists a feasible $(\tilde{m}, \tilde{w}) \in \mathcal{M}_h \times \mathcal{W}_h$ such that
$\tilde{w} \in \text{int}(K), \quad \tilde{m}_{i,j} > 0 \ \forall\, (i,j).$

Existence (BA-Kalise-Silva, 2018)

For any $\nu \geq 0$ we have

1. (P_h) admits at least one solution and the optimal costs are finite.
2. Let (m^h, w^h) be a solution to (P_h). Then, there exists $(u^h, \lambda^h) \in \mathcal{M}_h \times \mathbb{R}$ s.t. DSMFG holds ($w^h = m^h[\widehat{D_h u^h}]$).

Proof: Qualification condition implies the existence of Lagrange multipliers. DSMFG follows from FOC.

(P_h)'s structure

- Assume $f(x,\cdot)$ increasing ($F(x,\cdot)$ convex).
- $\varphi\colon (m,w) \mapsto \sum_{i,j} \phi_{i,j}(m_{i,j}, w_{i,j})$ where, $\forall\, 0 \le i,j \le N_h - 1$, $\phi_{i,j}(m_{i,j}, w_{i,j}) = \hat{b}(m_{i,j}, w_{i,j}) + F(x_{i,j}, m_{i,j})$ is proper, convex, l.s.c., non-smooth.
- Denote $-\nu\Delta_h = A$ and $\mathrm{div}_h = B$.

Reformulation of (P_h)

$$\min_{(m,w)\in\mathcal{M}_h\times\mathcal{W}_h} \varphi(m,w)$$
$$\text{s.t.} \quad \Xi(m,w) = (0,1),$$

where
$$\Xi = \begin{bmatrix} A & B \\ h^2\mathbf{1}^\top & 0 \end{bmatrix}.$$

(P_h)'s structure

Reformulation of (P_h)

$$\min_{(m,w)\in\mathcal{M}_h\times\mathcal{W}_h} \varphi(m,w)$$
$$\text{s.t.} \quad \Xi(m,w) = (0,1).$$

- (P_h) is equivalent to

$$(P_h) \qquad \min_{(m,w)\in\mathcal{M}_h\times\mathcal{W}_h} \Phi(m,w) = \varphi(m,w) + \psi(L(m,w))$$

$$\text{(Split)} \begin{cases} \psi = \iota_{\{(0,1)\}} \\ L = \Xi \end{cases} \qquad \text{or} \qquad \text{(Unsplit)} \begin{cases} \psi = \iota_{\Xi^{-1}(0,1)} \\ L = \text{Id}. \end{cases}$$

1. Motivation

2. (P_h)

3. Splitting algorithms: review

4. Split/unsplit approaches

5. Numerical experiences

General convex optimization problem

Problem (P)

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; \Phi(x),$$

assuming that solution set $Z \neq \varnothing$.

- $\Phi\colon \mathbb{R}^N \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ is
  - **convex :** $(\forall x, y \in \mathbb{R}^N)(\forall \lambda \in [0,1])$ $\Phi(x+\lambda(y-x)) \leq \Phi(x) + \lambda(\Phi(y) - \Phi(x))$.
  - **lower semicontinuous (l.s.c):** $(\forall x \in \mathbb{R}^N) \quad \Phi(x) \leq \liminf_{y\to x} \Phi(y)$.
  - **proper:** Φ is not always $+\infty$ and never $-\infty$.
- Functions satisfying above conditions constitute the class $\Gamma_0(\mathbb{R}^N)$.

Φ is smooth

Problem (P)

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \quad \Phi(x)$$

- Φ is convex differentiable and $\nabla\Phi$ is χ-Lipschitz. $\|\nabla\Phi(x) - \nabla\Phi(y)\| \leq \chi\|x - y\|$.

Optimality conditions

$$x^* \text{ is a solution to (P)} \quad \Leftrightarrow \quad 0 = \nabla\Phi(x^*).$$

- Equivalent to, for every $\gamma > 0$, $x^* = x^* - \gamma\nabla\Phi(x^*)$.
- Equivalent to, for every $\gamma > 0$, x^* is a fixed point of

Gradient operator

$$G_{\gamma\Phi}x = x - \gamma\nabla\Phi(x).$$

Φ is smooth

- We can approximate solutions to (P) via fixed points iterations of the operator $G_{\gamma\Phi}$, i.e.,

Gradient method

$$x_0 \in \mathbb{R}^N, \quad x_{n+1} = G_{\gamma\Phi} x_n = x_n - \gamma \nabla \Phi(x_n).$$

- When the operator $G_{\gamma\Phi}$ is good enough for obtaining convergence ?

Φ is smooth

- We can approximate solutions to (P) via fixed points iterations of the operator $G_{\gamma\Phi}$, i.e.,

Gradient method

$$x_0 \in \mathbb{R}^N, \quad x_{n+1} = G_{\gamma\Phi} x_n = x_n - \gamma \nabla \Phi(x_n).$$

- When the operator $G_{\gamma\Phi}$ is good enough for obtaining convergence ?
  - Φ $\beta-$strongly convex ($\Phi - \beta\|\cdot\|^2/2$ convex), $0 < \gamma < 2/\chi$:

Contraction

$$\|G_{\gamma\Phi}x - G_{\gamma\Phi}y\| \leq r\|x - y\|$$

  - $r = \max\{|1-\gamma\beta|, |1-\gamma\chi|\} < 1.$ $\gamma^* = 2/(\beta+\chi)$
  - Banach-Picard (linear convergence to unique solution).

Φ is smooth

- We can approximate solutions to (P) via fixed points iterations of the operator $G_{\gamma\Phi}$, i.e.,

Gradient method

$$x_0 \in \mathbb{R}^N, \quad x_{n+1} = G_{\gamma\Phi} x_n = x_n - \gamma \nabla \Phi(x_n).$$

- When the operator $G_{\gamma\Phi}$ is good enough for obtaining convergence ?
  - $0 < \gamma < 2/\chi$:

Averaged nonexpansive

$$\|G_{\gamma\Phi}x - G_{\gamma\Phi}y\|^2 \leq \|x-y\|^2 - \left(\frac{1-\mu}{\mu}\right) \|(\mathrm{Id} - G_{\gamma\Phi})x - (\mathrm{Id} - G_{\gamma\Phi})y\|^2$$

  - $\mu = \gamma\chi/2$
  - Convergence $O(1/k)$ to a solution.

Φ is smooth

- We can approximate solutions to (P) via fixed points iterations of the operator $G_{\gamma\Phi}$, i.e.,

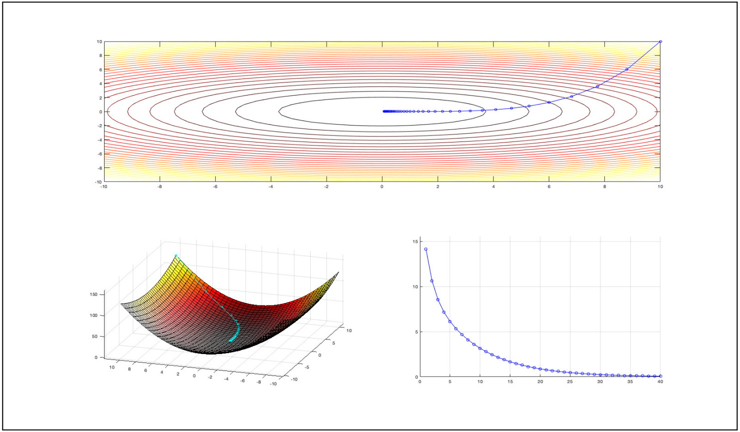
Gradient method

$$x_0 \in \mathbb{R}^N, \quad x_{n+1} = G_{\gamma\Phi} x_n = x_n - \gamma \nabla \Phi(x_n).$$

- When the operator $G_{\gamma\Phi}$ is good enough for obtaining convergence ?
  - Variants with γ_k yields $O(1/k^2)$. (Nesterov, 1983)

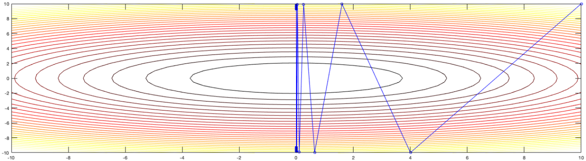
Example Grad: $\gamma = 0.2 \in]0, 2/\chi[$

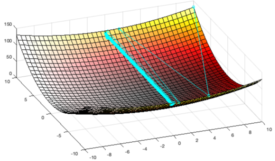
- $\Phi\colon (x,y) \mapsto 0.3\,x^2 + y^2$
- $\nabla\Phi\colon (x,y) \mapsto (0.6\,x, 2\,y)$
- $\chi = 2$, $\beta = 0.6$

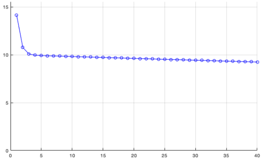


Example Grad: $\gamma = 0.999 \in \,]0, 2/\chi[$

- $\Phi\colon (x,y) \mapsto 0.3\,x^2 + y^2$
- $\nabla\Phi\colon (x,y) \mapsto (0.6\,x, 2\,y)$
- $\chi = 2,\ \beta = 0.6$

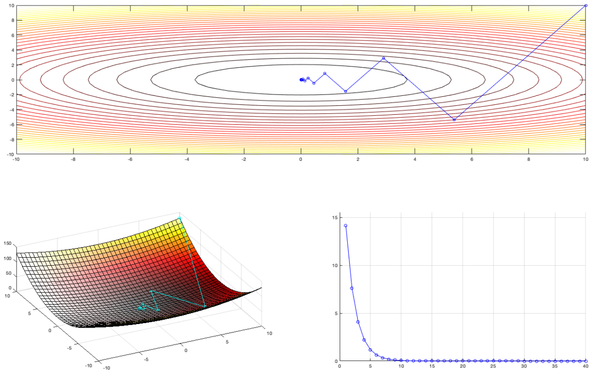






Example Grad: $\gamma = \gamma^* \in \left]0, 2/\chi\right[$

- $\Phi\colon (x,y) \mapsto 0.3\,x^2 + y^2$
- $\nabla\Phi\colon (x,y) \mapsto (0.6\,x, 2\,y)$
- $\chi = 2,\ \beta = 0.6,\ \gamma^* = 2/(\beta+\chi)$

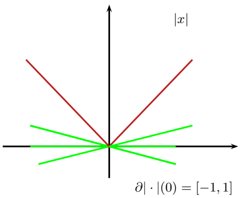


$\Phi \in \Gamma_0(\mathbb{R}^N)$ is non-smooth

Problem (P)

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; \Phi(x)$$

$$\partial\Phi\colon x \mapsto \left\{u \in \mathbb{R}^N \;\middle|\; (\forall y \in \mathbb{R}^N)\;\; \Phi(x) + u^\top(y-x) \leq \Phi(y)\right\}$$



Optimality conditions

$$x^* \text{ is a solution to (P)} \quad \Leftrightarrow \quad 0 \in \partial\Phi(x^*).$$

$\Phi \in \Gamma_0(\mathbb{R}^N)$ is non-smooth

Problem (P)

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; \Phi(x)$$

Optimality conditions

$$x^* \text{ is a solution to (P)} \quad \Leftrightarrow \quad 0 \in \partial\Phi(x^*).$$

- Equivalent to, for every $\gamma > 0$, $x^* \in x^* + \gamma\partial\Phi(x^*)$.
- Equivalent to, for every $\gamma > 0$, x^* is a fixed point of

Proximity operator

$$\text{prox}_{\gamma\Phi} = (\text{Id} + \gamma\partial\Phi)^{-1} : x \mapsto \underset{y\in\mathbb{R}^N}{\text{argmin}} \;\; \gamma\Phi(y) + \frac{1}{2}\|y - x\|^2.$$

- $\text{prox}_{\iota_C} = P_C$.

$\Phi \in \Gamma_0(\mathbb{R}^N)$ is non-smooth

- We can approximate solutions to (P) via fixed points iterations of the operator $\text{prox}_{\gamma\Phi}$, i.e.,

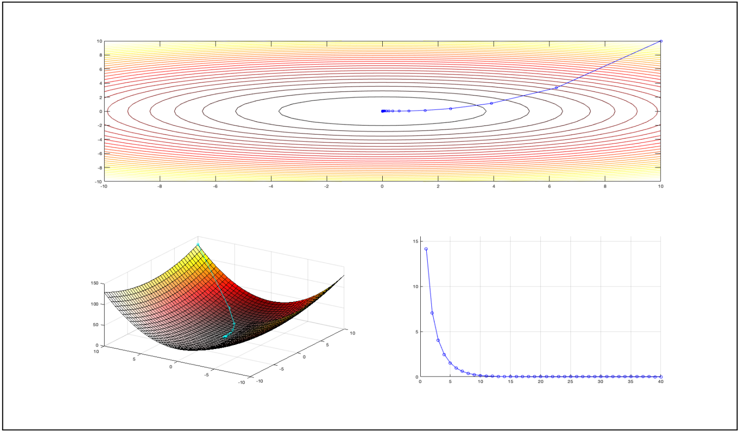
Proximal point algorithm (PPA)

$$x_0 \in \mathbb{R}^N, \quad x_{n+1} = \text{prox}_{\gamma\Phi}(x_n).$$

- When the operator $\text{prox}_{\gamma\Phi}$ is good enough ? Always (for every $\gamma > 0$) (Martinet 1970-72; Rockafellar 1976).
  - Φ $\beta-$strongly convex: $\text{prox}_{\gamma\Phi}$ contraction $(r = \frac{1}{1+\gamma\beta} < 1)$.
  - In general, it is averaged nonexpansive $(\mu = \frac{1}{2})$. $O(1/k)$.
  - Variations considering γ_k yields $O(1/k^2)$ (Güler 1992).

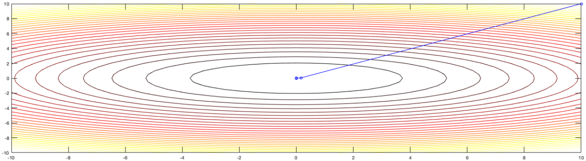
Example PPA: $\gamma = 1$

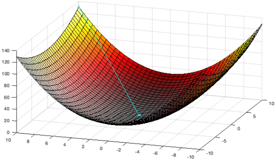
- $\Phi\colon (x,y) \mapsto 0.3\,x^2 + y^2$
- $\text{prox}_{\gamma\Phi}\colon (x,y) \mapsto (x/(1+0.6\,\gamma), y/(1+2\,\gamma))$
- $\chi = 2,\ \beta = 0.6$

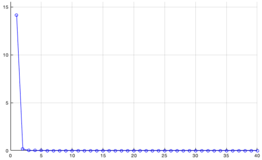


Example PPA: $\gamma = 100$

- $\Phi\colon (x,y) \mapsto 0.3\,x^2 + y^2$
- $\mathrm{prox}_{\gamma\Phi}\colon (x,y) \mapsto (x/(1+0.6\,\gamma), y/(1+2\,\gamma))$
- $\chi = 2,\ \beta = 0.6$







Our case: $\Phi = \varphi + \psi \circ L$

Problem (P)

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \quad \Phi(x) = \varphi(x) + \psi(Lx)$$

- **Lagrangian approach:** Solve

$$\min_{Lx=y} \varphi(x) + \psi(y)$$

Augmented Lagrangian

$$\mathcal{L}_c(x, y, u) = \varphi(x) + \psi(y) + u^\top (Lx - y) + \frac{c}{2} \|Lx - y\|^2$$

Our case: $\Phi = \varphi + \psi \circ L$

- Under qualification conditions x solves (P) iff (x, Lx, u) is a saddle point of $\mathcal{L}_c$, with $c > 0$.
- Alternating minimization-maximization of $\mathcal{L}_c$ we obtain:

ADMM (Glowinski-Marrocco 1975; Gabay-Mercier 1976)

$$
\begin{aligned}
x^{k+1} &= \text{argmin}_x \mathcal{L}_c(x, y^k, u^k) \\
&= \text{argmin}_x \left\{ \varphi(x) + u^{k\top} Lx + \frac{\gamma}{2} \|Lx - y^k\|^2 \right\}
\end{aligned}
$$

$$
y^{k+1} = \text{argmin}_y \mathcal{L}_c(x^{k+1}, y, u^k) = \text{prox}_{\psi/\gamma}(u^k/\gamma + Lx^{k+1})
$$

$$
u^{k+1} = u^k + \gamma(Lx^{k+1} - y^{k+1}).
$$

- Problem: The first step is not easy to compute in general.
- Predictor-corrector proximal multiplier (PCPM) splits with additional multiplier update (Chen-Teboulle, 1994)

Our case: $\Phi = \varphi + \psi \circ L$

Problem (P)

$$\min_{x\in\mathbb{R}^N} \; \Phi(x) = \varphi(x) + \psi(Lx)$$

Fenchel-Rockafellar duality yields:

Problem (D)

$$\min_{u\in\mathbb{R}^M} \; \varphi^*(-L^\top u) + \psi^*(u)$$

Fenchel conjugate

$$\psi^* \colon u \mapsto \sup_{v\in\mathbb{R}^M} \left(u^\top v - \psi(v)\right)$$

- $\psi \in \Gamma_0(\mathbb{R}^M) \Leftrightarrow \psi^* \in \Gamma_0(\mathbb{R}^M)$
- $\partial\psi^* = (\partial\psi)^{-1}$

Our case: $\Phi = \varphi + \psi \circ L$

Under qualification conditions, we have

$$
\begin{aligned}
x \text{ solves } (P) \quad &\Leftrightarrow \quad 0 \in \partial\varphi(x) + L^\top \partial\psi(Lx) \\
&\Leftrightarrow \quad (\exists\, u \in \partial\psi(Lx)) \quad 0 \in \partial\varphi(x) + L^\top u \\
&\Leftrightarrow \quad (\exists\, u \in \mathbb{R}^M) \quad \begin{cases} Lx \in (\partial\psi)^{-1} u = \partial\psi^*(u) \\ 0 \in \partial\varphi(x) + L^\top u \end{cases} \\
&\Leftrightarrow \quad (\exists\, u \in \mathbb{R}^M) \quad \begin{cases} 0 \in \partial\psi^*(u) - Lx \\ 0 \in \partial\varphi(x) + L^\top u \end{cases} \\
&\Leftrightarrow \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \underbrace{\begin{bmatrix} \partial\varphi & 0 \\ 0 & \partial\psi^* \end{bmatrix}}_{\boldsymbol{M}} \underbrace{\begin{bmatrix} x \\ u \end{bmatrix}}_{\boldsymbol{X}} + \underbrace{\begin{bmatrix} 0 & L^\top \\ -L & 0 \end{bmatrix}}_{\boldsymbol{S}} \underbrace{\begin{bmatrix} x \\ u \end{bmatrix}}_{\boldsymbol{X}}
\end{aligned}
$$

Our case: $\Phi = \varphi + \psi \circ L$

$$x \text{ solves } (P) \quad \Leftrightarrow \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \underbrace{\begin{bmatrix} \partial\varphi & 0 \\ 0 & \partial\psi^* \end{bmatrix}}_{\boldsymbol{M}} \underbrace{\begin{bmatrix} x \\ u \end{bmatrix}}_{\boldsymbol{X}} + \underbrace{\begin{bmatrix} 0 & L^\top \\ -L & 0 \end{bmatrix}}_{\boldsymbol{S}} \underbrace{\begin{bmatrix} x \\ u \end{bmatrix}}_{\boldsymbol{X}}$$

- for some u solving (D).

Our case: $\Phi = \varphi + \psi \circ L$

$$x \text{ solves } (P) \quad \Leftrightarrow \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \underbrace{\begin{bmatrix} \partial\varphi & 0 \\ 0 & \partial\psi^* \end{bmatrix}}_{\boldsymbol{M}} \underbrace{\begin{bmatrix} x \\ u \end{bmatrix}}_{\boldsymbol{X}} + \underbrace{\begin{bmatrix} 0 & L^\top \\ -L & 0 \end{bmatrix}}_{\boldsymbol{S}} \underbrace{\begin{bmatrix} x \\ u \end{bmatrix}}_{\boldsymbol{X}}$$

- for some u solving (D).
- $\boldsymbol{M}(x,u) = \partial\boldsymbol{\Psi}(x,u)$, where $\boldsymbol{\Psi}(x,u) = \varphi(x) + \psi^*(u)$. ✓

Our case: $\Phi = \varphi + \psi \circ L$

$$x \text{ solves } (P) \quad \Leftrightarrow \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \underbrace{\begin{bmatrix} \partial\varphi & 0 \\ 0 & \partial\psi^* \end{bmatrix}}_{\boldsymbol{M}} \underbrace{\begin{bmatrix} x \\ u \end{bmatrix}}_{\boldsymbol{X}} + \underbrace{\begin{bmatrix} 0 & L^\top \\ -L & 0 \end{bmatrix}}_{\boldsymbol{S}} \underbrace{\begin{bmatrix} x \\ u \end{bmatrix}}_{\boldsymbol{X}}$$

- for some u solving (D).
- $\boldsymbol{M}(x,u) = \partial\boldsymbol{\Psi}(x,u)$, where $\boldsymbol{\Psi}(x,u) = \varphi(x) + \psi^*(u)$. ✓
- $\boldsymbol{S} \neq \nabla \boldsymbol{G}$... but it is maximally monotone: $(\boldsymbol{S}\boldsymbol{X} - \boldsymbol{S}\boldsymbol{Y})^\top(\boldsymbol{X} - \boldsymbol{Y}) \geq 0$ and continuous.

For a monotone operator $\boldsymbol{S}$, we have:

Resolvent

$$J_{\gamma\boldsymbol{S}} = (\mathrm{Id} + \gamma\boldsymbol{S})^{-1}$$

Explicit operator

$$G_{\gamma\boldsymbol{S}} = \mathrm{Id} - \gamma\boldsymbol{S}$$

Our case: $\Phi = \varphi + \psi \circ L$

Splitting approach:

- $J_{\gamma \boldsymbol{M}} = \operatorname{prox}_{\gamma \boldsymbol{\Psi}} \colon (x,u) \mapsto (\operatorname{prox}_{\gamma\varphi} x, \operatorname{prox}_{\gamma\psi^*} u)$
- $G_{\gamma \boldsymbol{S}} = \operatorname{Id} - \gamma \boldsymbol{S} \colon (x,u) \mapsto (x - \gamma L^\top u, u + \gamma L x)$

Mon.+Skew (BA-Combettes 2011)

Let $0 < \gamma < \|L\|^{-1}$, $x_0 \in \mathbb{R}^N$ and $u_0 \in \mathbb{R}^M$ and iterate

$$\left\lfloor \begin{aligned} p_{1,n} &= \operatorname{prox}_{\gamma\varphi}(x_n - \gamma L^\top u_n) \\ p_{2,n} &= \operatorname{prox}_{\gamma\psi^*}(u_n + \gamma L x_n) \\ x_{n+1} &= p_{1,n} - \gamma(L^\top p_{2,n} - L^\top u_n) \\ u_{n+1} &= p_{2,n} + \gamma(L p_{1,n} - L x_n). \end{aligned} \right.$$

Then $(x_n, u_n) \to (\bar{x}, \bar{u})$, where $\bar{x}$ solves (P) and $\bar{u}$ solves (D).

- If φ & ψ^* strongly convex: linear conv. (Tseng 2000).

Our case: $\Phi = \varphi + \psi \circ L$

Problem (P)

$$\min_{x\in\mathbb{R}^N} \; \Phi(x) = \varphi(x) + \psi(Lx)$$

Equivalent to (Condat-Vũ, 2013 & He-Yuan, 2012)

$$\begin{bmatrix}0\\0\end{bmatrix} \in \underbrace{\left(\begin{bmatrix}\partial\varphi & 0\\ 0 & \partial\psi^*\end{bmatrix} + \begin{bmatrix}0 & L^\top\\ -L & 0\end{bmatrix}\right)}_{\boldsymbol{A}=\boldsymbol{M}+\boldsymbol{S}} \underbrace{\begin{bmatrix}x\\u\end{bmatrix}}_{\boldsymbol{X}}$$

- $J_{\boldsymbol{A}}$ is not explicit !

Our case: $\Phi = \varphi + \psi \circ L$

Problem (P)

$$\min_{x \in \mathbb{R}^N} \Phi(x) = \varphi(x) + \psi(Lx)$$

Equivalent to (Condat-Vũ, 2013 & He-Yuan, 2012)

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \underbrace{\left(\begin{bmatrix} \partial\varphi & 0 \\ 0 & \partial\psi^* \end{bmatrix} + \begin{bmatrix} 0 & L^\top \\ -L & 0 \end{bmatrix} \right)}_{\boldsymbol{A}=\boldsymbol{M}+\boldsymbol{S}} \underbrace{\begin{bmatrix} x \\ u \end{bmatrix}}_{\boldsymbol{X}}$$

- $J_{\boldsymbol{A}}$ is not explicit !
- Set $\boldsymbol{U} \succ 0$ symmetric and $\langle \boldsymbol{X} \mid \boldsymbol{Y} \rangle_{\boldsymbol{U}} = \boldsymbol{X}^\top \boldsymbol{U} \boldsymbol{Y}$.
- $\langle \boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{X} - \boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{Y} \mid \boldsymbol{X} - \boldsymbol{Y} \rangle_{\boldsymbol{U}} = (\boldsymbol{A}\boldsymbol{X} - \boldsymbol{A}\boldsymbol{Y})^\top (\boldsymbol{X} - \boldsymbol{Y})$.
- $\boldsymbol{U}^{-1}\boldsymbol{A}$ is monotone under this metric: PPA.

Our case: $\Phi = \varphi + \psi \circ L$

Specific $\boldsymbol{U}$ allows $J_{\boldsymbol{U}^{-1}\boldsymbol{A}}$ explicit:

Chambolle-Pock (2011)

$x_0, \bar{x}_0 \in \mathbb{R}^N$ and $u_0 \in \mathbb{R}^M$, $\tau, \gamma > 0$ such that $\tau\gamma\|L\|^2 < 1$

$$\left\lfloor \begin{array}{l} x^{n+1} = \operatorname{prox}_{\tau\varphi}(x^n - \tau L^\top u^n) \\ u^{n+1} = \operatorname{prox}_{\gamma\psi^*}(u^n + \gamma L(2x^{n+1} - x^n)). \end{array} \right.$$

Acceleration (Chambolle-Pock 2011):

- φ and ψ^* strongly convex: linear convergence
- φ or ψ^* strongly convex and τ_k, γ_k: $\|x_k - x^*\| \leq C/k$.

1. Motivation
2. (P_h)
3. Splitting algorithms: review
4. Split/unsplit approaches
5. Numerical experiences

(P_h)'s structure

Reformulation of (P_h)

$$\min_{(m,w)\in\mathcal{M}_h\times\mathcal{W}_h} \varphi(m,w)$$
$$\text{s.t.}\quad \Xi(m,w)=(0,1),$$

$$(P_h)\qquad \min_{(m,w)\in\mathcal{M}_h\times\mathcal{W}_h} \Phi(m,w)=\varphi(m,w)+\psi(L(m,w))$$

$$(\text{Split})\begin{cases}\psi=\iota_{\{(0,1)\}}\\ L=\Xi\end{cases}\qquad\text{or}\qquad(\text{Unsplit})\begin{cases}\psi=\iota_{\Xi^{-1}(0,1)}\\ L=\text{Id}.\end{cases}$$

All previous methods implement the dual variable update

$$u^{n+1}=\text{prox}_{\gamma\psi^*}(u^n+\gamma L\widetilde{x}^n).$$

Split/unsplit approaches

“Split” decomposition:

- $\psi = \iota_{\{(0,1)\}}$, $\text{prox}_{\gamma\psi^*} = \text{Id} - \gamma(0,1)$, and $L = \Xi$.
- Then, all previous methods include a Lagrange multiplier step of the form

$$u^{k+1} = u^k + \gamma(\Xi x^k - (0,1))$$

- The primal iterates $(x_k)_{k\in\mathbb{N}}$ are not feasible !
- Very slow...

“Unsplit” decomposition:

- $\psi = \iota_{\{\Xi^{-1}(0,1)\}}$ and for computing $\text{prox}_{\gamma\psi^*}$ or $\text{prox}_{\gamma\psi}$ we need to invert $\Xi\Xi^*$ (more precisely $\nu^2 AA^* + BB^*$).
- Depending on the parameter ν, this matrix can be very bad conditioned and difficult to invert.

Projected Chambolle-Pock splitting[5]

We avoid matrix inversions along with ensuring primal iterates to satisfy some of the constraints.

Projected Chambolle-Pock (PCP)

Let $x_0 \in \mathcal{H}$, $u_0 \in \mathcal{G}$ and $\sigma\tau\|L\|^2 < 1$.

$$(\forall n \in \mathbb{N}) \quad \begin{cases} p_{n+1} = \mathrm{prox}_{\tau\varphi}(x_n - \tau L^* u_n) \\ x_{n+1} = P_C p_{n+1} \\ u_{n+1} = \mathrm{prox}_{\sigma\psi^*}(u_n + \sigma L(x_{n+1} + p_{n+1} - x_n)). \end{cases}$$

$(x^k)_{k\in\mathbb{N}} \subset C$ converges to a point in $Z \cap C = Z$.

- In particular, we use C as the mass constraint.
- C can change deterministically/randomly among iterations.

[5] with J. Deride, S. López Rivera, and C. Vega

$\text{prox}_{\gamma\varphi}$ computation

- In all previous methods we need to compute $\text{prox}_{\gamma\varphi}(m, w)$.
- Recall that $\varphi\colon (m, w) \mapsto \sum_{i,j} \phi_{i,j}(m_{i,j}, w_{i,j})$, where $\phi_{i,j}\colon (\mu, \omega) \mapsto \hat{b}(\mu, \omega) + F_{i,j}(\mu)$ and $F_{i,j} = F(x_{i,j}, \cdot)$.
- We have $\text{prox}_{\gamma\varphi}(m, w) = (\text{prox}_{\gamma\phi_{i,j}}(m_{i,j}, w_{i,j}))_{i,j}$.

Prox computation

$$\text{prox}_{\gamma\phi}\colon (\mu, \omega) \mapsto \begin{cases} (0, 0), & \text{if } m \leq \gamma F'(0); \\ (p^*, p^*\, P_K w/(p^* + \gamma)), & \text{if } m > \gamma F'(0), \end{cases}$$

where $p^* \geq 0$ is the unique solution to

$$(p + \gamma F'(p) - m)(p + \gamma)^2 - \frac{\gamma}{2}|P_K w|^2 = 0.$$

- We extend prox in Papadakis, Peyre & Oudet (2014) used in the context of optimal transport (we include F and K).

Algorithm

Denoting $\varphi = b_2 + F$, $A = -\nu\Delta_h$, $B = \text{div}_h$, the classic CP splitting reads :

$$\begin{pmatrix} u^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} u^k + \gamma(A\bar{m}^k + B\bar{w}^k) \\ \lambda^k + \gamma(h^2 \sum_{i,j} \bar{m}^k_{i,j} - 1) \end{pmatrix}$$

$$\begin{pmatrix} m^{k+1} \\ w^{k+1} \end{pmatrix} = \text{prox}_{\tau\varphi} \begin{pmatrix} m^k - \tau(A^\top u^{k+1} + h^2\lambda^{k+1}\mathbf{1}) \\ w^k - \tau B^\top u^{k+1} \end{pmatrix}$$

$$\begin{pmatrix} \bar{m}^{k+1} \\ \bar{w}^{k+1} \end{pmatrix} = \begin{pmatrix} 2m^{k+1} - m^k \\ 2w^{k+1} - w^k \end{pmatrix}$$

but primal iterates do not satisfy any of the constraints, which extremely affect the speed of the method.

Algorithm

Imposing, the constraint $\int m = 1$ the PCP reads

$$\begin{pmatrix} u^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} u^k + \gamma(A\bar{m}^k + B\bar{w}^k) \\ \lambda^k + \gamma(h^2 \sum_{i,j} \bar{m}^k_{i,j} - 1) \end{pmatrix}$$

$$\begin{pmatrix} n^{k+1} \\ v^{k+1} \end{pmatrix} = \operatorname{prox}_{\tau\varphi} \begin{pmatrix} m^k - \tau(A^\top u^{k+1} + h^2 \lambda^{k+1} \mathbf{1}) \\ w^k - \tau B^\top u^{k+1} \end{pmatrix}$$

$$\begin{pmatrix} m^{k+1} \\ w^{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{1} + \left(n^{k+1} - \mathbf{1} \sum_{i,j=1}^{N_h} n^{k+1}_{i,j}\right) \\ v^{k+1} \end{pmatrix}$$

$$\begin{pmatrix} \bar{m}^{k+1} \\ \bar{w}^{k+1} \end{pmatrix} = \begin{pmatrix} m^{k+1} + n^{k+1} - m^k \\ w^{k+1} + v^{k+1} - w^k \end{pmatrix}$$

which is much faster, and any of the primal iterates satisfy the imposed constraint. By including splitting, we do not need any matrix inversion.

Test 1

We consider the first-order stationary MFG system (Almulla-Ferreira-Gomes, 2015)

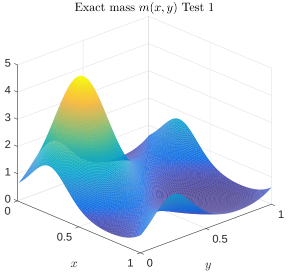
$$\frac{1}{2}|\nabla u|^2 - \lambda = \log m - \sin(2\pi x) - \sin(2\pi y),$$

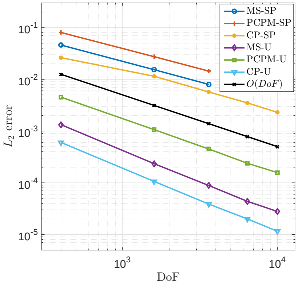
$$\operatorname{div}(m\nabla u) = 0, \quad \int_{\mathbb{T}^2} m dx = 1, \quad \int_{\mathbb{T}^2} u dx = 0,$$

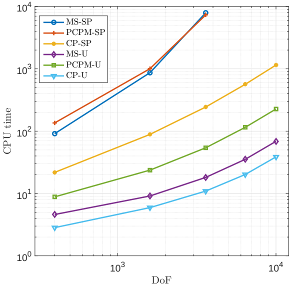
with explicit solution

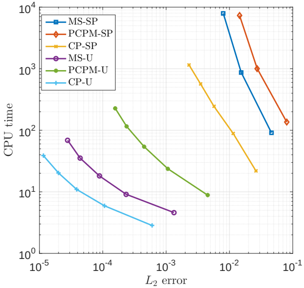
$$u(x,y) = 0, \quad m(x,y) = e^{\sin(2\pi x)+\sin(2\pi y)-\lambda},$$

$$\lambda = \log\left(\int_{\mathbb{T}^2} e^{\sin(2\pi x)+\sin(2\pi y)} dxdy\right).$$


Exact mass $m(x, y)$ Test 1
5
4
3
2
1
0
0
0.5
1
x
0
0.5
1
y


MS-SP
PCPM-SP
CP-SP
MS-U
PCPM-U
CP-U
$O(DoF)$
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
L_2 error
10^{3}
10^{4}
DoF

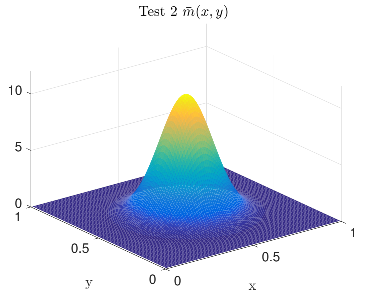

MS-SP
PCPM-SP
CP-SP
MS-U
PCPM-U
CP-U
10^{4}
10^{3}
10^{2}
10^{1}
10^{0}
CPU time
10^{3}
10^{4}
DoF

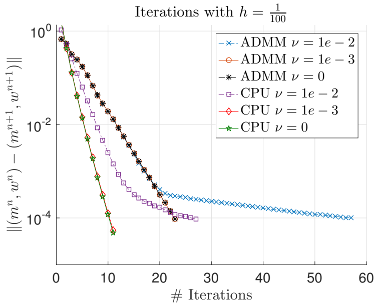

MS-SP
PCPM-SP
CP-SP
MS-U
PCPM-U
CP-U
10^{4}
10^{3}
10^{2}
10^{1}
10^{0}
CPU time
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
L_2 error

Test 2: Comparison with ADMM

$$f(x,y,m) = \frac{1}{2}(m - \bar{m})$$

where $\bar{m}$ is centered gaussian density.





Concluding remarks

Also done:

- We tackle the more general case $H\colon (x,p) \mapsto |p|^{q'}/q'$ for $q > 1$.
- We also present 2 additional tests.
  - Test 3: We include hard density constraints in some regions.
  - Test 4: We vary $q \neq 2$.
- Dynamic case.

In preparation:

- More general Hamiltonians, congestion, non-local couplings.
- Other efficient projection configurations for avoiding matrix inversions.

References

- Y. Achdou, I. Capuzzo Dolcetta, Mean field games: Numerical methods. *SIAM J. Numer. Anal.*, 48:1136–1162, 2010.
- Y. Achdou, F. Camilli, I. Capuzzo Dolcetta. Mean field games: convergence of a finite difference method. *SIAM J. Numer. Anal.*, 51:2585–2612, 2013.
- L. M. Briceño-Arias, S. López Rivera, A projected primal-dual splitting for solving constrained monotone inclusions, *J. Optim. Theory Appl.*, 180:907–924, 2019.
- L. M. Briceño-Arias, P. L. Combettes, A monotone+skew splitting model for composite monotone inclusions in duality, *SIAM J. Optim.*, 21:1230–1250, 2011.
- L. M. Briceño-Arias, D. Kalise, F. J. Silva, Proximal methods for stationary Mean Field Games with local couplings, *SIAM J. Control Optim.*, 56: 801–836, 2018.
- L.M. Briceño-Arias, D. Kalise, Z. Kobeissi, M. Laurière, A. Mateos González, F. J. Silva, On the implementation of a primal-dual algorithm for second order time-dependent mean field games with local couplings, *ESAIM Proc. Surveys*, 65:330–348, 2019.
- A. Chambolle, T. Pock. A first order primal dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.*, 40:120–145, 2011.
- G. Chen, M. Teboulle. A proximal-based decomposition method for convex minimization problems. *Math. Programming Ser. A*, 64:81–101, 1994.

References

- M. Huang, R. P. Malhamé, P. E. Caines. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Commun. Inf. Syst.* 6:221–251, 2006.
- J.-M. Lasry and P.-L. Lions. Mean field games. *Jpn. J. Math.* 2:229–260, 2007.

References

- M. Huang, R. P. Malhamé, P. E. Caines. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Commun. Inf. Syst.* 6:221–251, 2006.
- J.-M. Lasry and P.-L. Lions. Mean field games. *Jpn. J. Math.* 2:229–260, 2007.

Merci de votre attention !