Optimisation et Aléa : un pas de deux

Gersende Fort

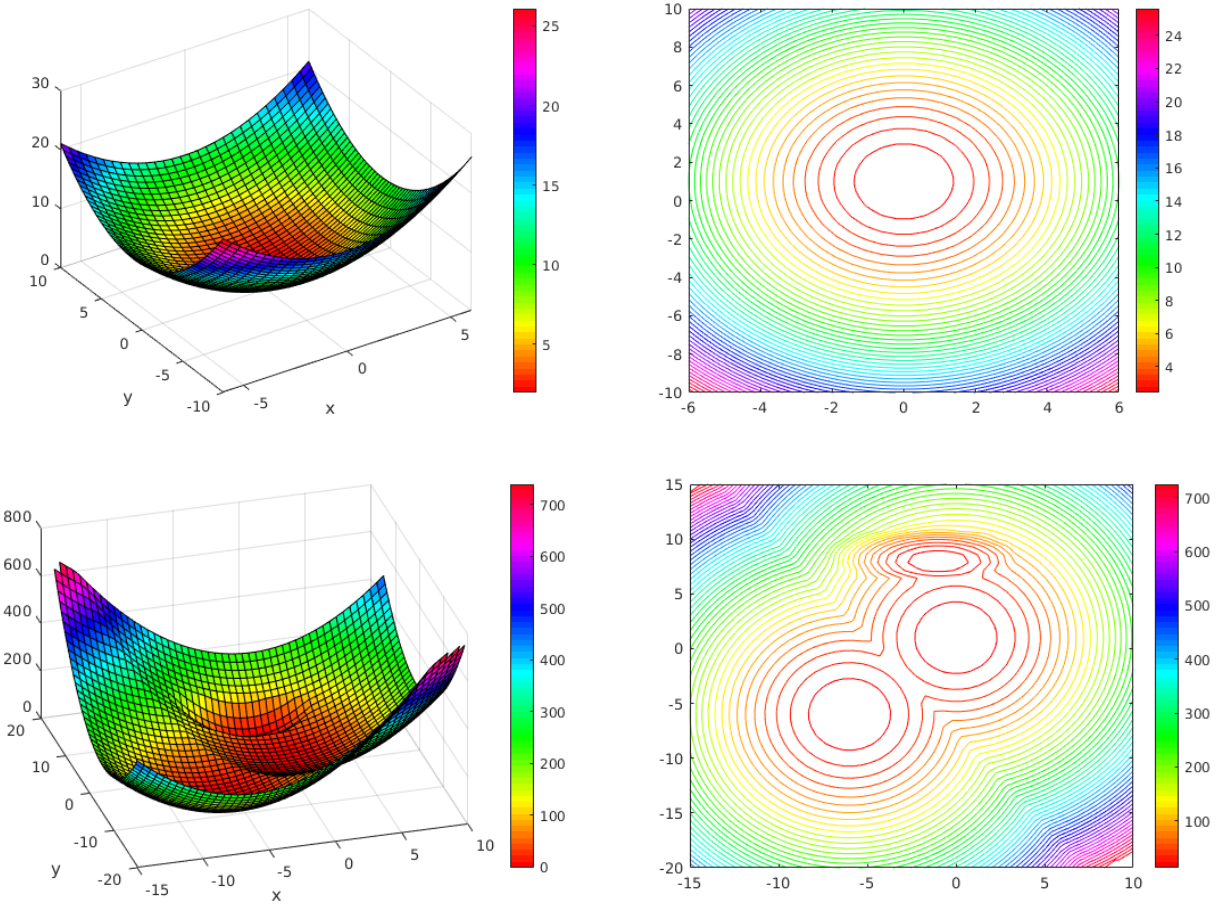
CNRS
Institut de Mathématiques de Toulouse

Mois de l'Optimisation, Toulouse, Nov. 2022

Entrelacement de l'optimisation et de l'aléatoire : optimisation

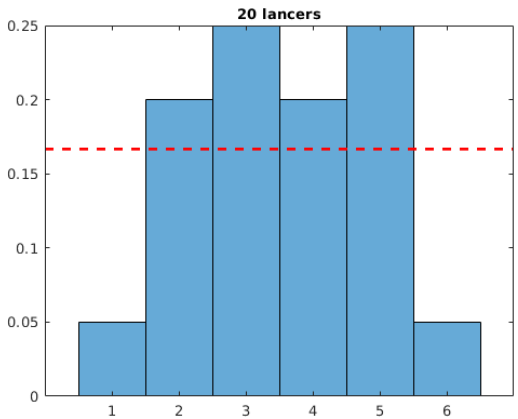
- une *fonction objectif* F, un *ensemble de contraintes* $\mathcal{S}$
- recherche d'une *configuration optimale* x_{opt}

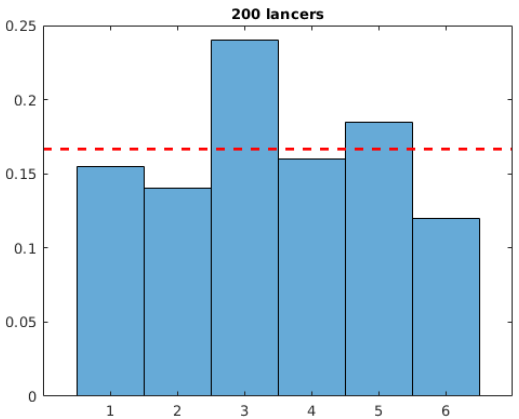
$$x_{\text{opt}} \in \mathcal{S} \quad \text{tel que} \quad F(x) \geq F(x_{\text{opt}}) \qquad \text{pour tout} \quad x \in \mathcal{S}.$$

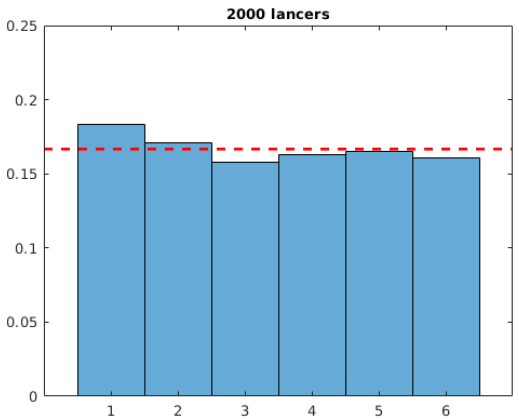


Entrelacement de l'optimisation et de l'aléatoire : l'aléatoire

- Aléatoire : Soumis au hasard, dont le résultat est incertain.
- Ex. d'expérience aléatoire : lancer de dés à six faces







Pour tout $j = 1, \cdots, 6$,

$$\lim_{N \to +\infty} \frac{1}{N} \sum_{k=1}^{N} 1_{\{j\}}(X_k) = \frac{1}{6}$$

Entrelacement de l'optimisation et de l'aléatoire : entrelacement

- Dans cette présentation, entrelacement illustré par la résolution de problèmes inverses par la Statistique
- l'optimisation rendue possible grâce à l'aléatoire / facilitée par l'aléatoire
- l'exploration aléatoire est améliorée grâce aux algorithmes d'optimisation.

Optimisation
Echantillonnage Monte Carlo

Quelques outils probabilistes
Loi des grands nombres
Autres résultats de Probabilités

L'aléa pour l'optimisation
Exemples d'algorithmes d'optimisation stochastique
Gradient stochastique
Expectation-Maximisation stochastique
Approximation Stochastique
Le SLAM

L'optimisation pour l'aléatoire : simulation Monte Carlo
Méthodes de Monte Carlo
Taux de reproduction de la Covid19

Conclusion

La loi des grands nombres

Soit des variables aléatoires $X_1, X_2, \cdots$, indépendantes, à valeur dans X et de même loi de probabilité ν.
Pour toute fonction f ν-intégrable

$$\text{p.s.} \qquad \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} f(X_k) = \int_{\mathsf{X}} f(x)\, \nu(\mathrm{d}x) \qquad \text{noté } \mathbb{E}\left[f(X)\right]$$

Un exemple de loi discrète

Le résultat de l'expérience est à valeur dans X au plus dénombrable

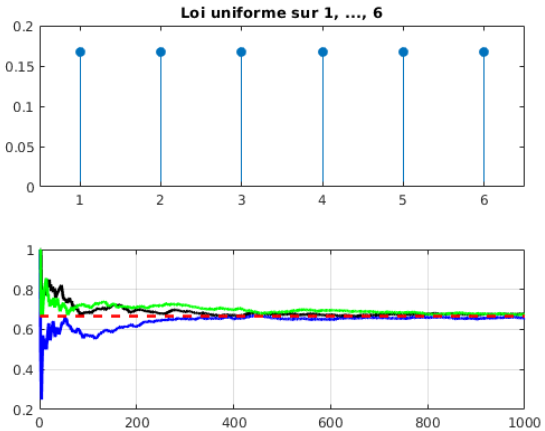
- Ex. Le lancer de dés : $\mathsf{X} = \{1, \cdots, 6\}$

Evolution de

$$N \mapsto \frac{1}{N} \sum_{k=1}^{N} 1_{]0,4]}(X_k)$$

le long d'une trajectoire de longueur $N_{\max} = 1\,000$.
La valeur limite est

$$\sum_{x \in \mathsf{X}} f(x)\, p(x) = \sum_{x=1}^{6} 1_{]0,4]}(x)\, \frac{1}{6} = \frac{2}{3}.$$



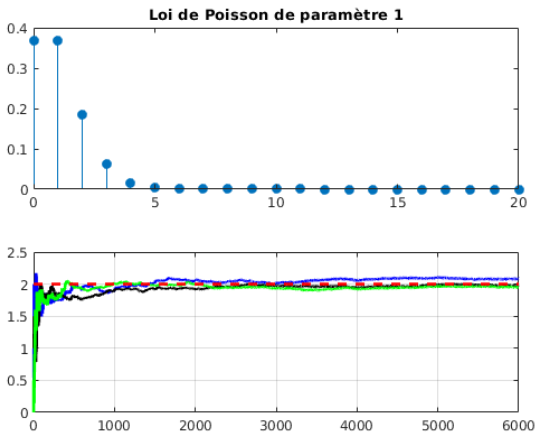
- Ex. Les données de comptes : $\mathsf{X} = \mathbb{N} = \{0, 1, \cdots\}$

Evolution de

$$N \mapsto \frac{1}{N} \sum_{k=1}^{N} X_k^2$$

le long d'une trajectoire de longueur $N_{\max} = 6\,000$.
La valeur limite est

$$\sum_{x=0}^{+\infty} x^2\, \frac{\lambda^x}{x!} \exp(-\lambda) = \lambda + \lambda^2.$$



Un exemple de loi continue

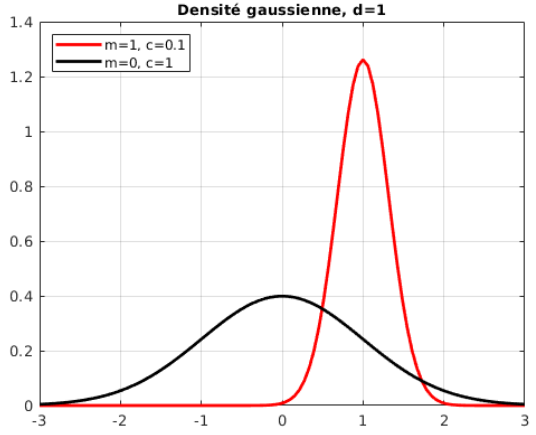
Le résultat de l'expérience est à valeur dans $\mathbb{R}^d$

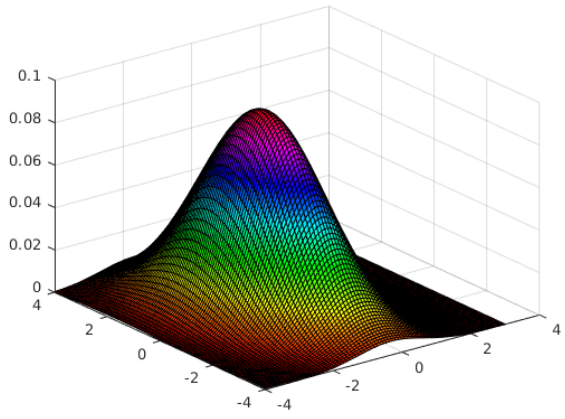
- Ex. Loi sur $\mathbb{R}$ de densité $p(x)$ par rapport à la mesure de Lebesgue

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{k=1}^{N} f(X_k) = \int_{-\infty}^{+\infty} f(x)\,p(x)\,\mathrm{d}x.$$

- Ex. Loi gaussienne sur $\mathbb{R}^d$ centrée en m et de matrice de covariance C

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{k=1}^{N} f(X_k) = \int_{\mathbb{R}^d} f(x)\,\frac{1}{\sqrt{2\pi}^d}\,\frac{\exp\left(-0.5(x-m)^\top C^{-1}(x-m)\right)}{\sqrt{\det(C)}}\,\mathrm{d}x.$$





La loi des grands nombres

Soit des variables aléatoires $X_1, X_2, \cdots$, indépendantes, à valeur dans X et de même loi de probabilité ν.
Pour toute fonction f ν-intégrable

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} f(X_k) = \int_{\mathsf{X}} f(x)\, \nu(\mathrm{d}x) \qquad \text{noté } \mathbb{E}\left[f(X)\right]$$

Dans les deux sens !

- Comportement moyen d'une suite de résultats $\{f(x_1), f(x_2), \cdots\}$ d'expérience aléatoire répétée à l'identique
- Une intégrale peut se relire comme la limite d'une somme partielle normalisée.

Quelques outils probabilistes
Loi des grands nombres
Autres résultats de Probabilités

L'aléa pour l'optimisation
Exemples d'algorithmes d'optimisation stochastique
Gradient stochastique
Expectation-Maximisation stochastique
Approximation Stochastique
Le SLAM

L'optimisation pour l'aléatoire : simulation Monte Carlo
Méthodes de Monte Carlo
Taux de reproduction de la Covid19

Conclusion

Autres résultats de Probabilités

- Théorème de la limite centrale

$$\sqrt{N}\left(\frac{1}{N}\sum_{k=1}^{N} f(X_k) - \mathbb{E}[f(X)]\right) \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0, C_{d,f}\right)$$

- Erreurs L^p, $\qquad f : \mathsf{X} \to \mathbb{R}$

$$\mathbb{E}\left[\left|\ \frac{1}{N}\sum_{k=1}^{N} f(X_k) - \mathbb{E}[f(X)]\ \right|^2\right] \leq \frac{M_{d,f}}{N}$$

- Inégalités de concentration $\qquad f : \mathsf{X} \to \mathbb{R}$

$$\mathbb{P}\left(\frac{1}{N}\sum_{k=1}^{N} f(X_k) - \mathbb{E}[f(X)] \geq \epsilon\right) \leq \cdots$$

- Martingale; par ex., lorsque échantillonnage Monte Carlo exact
- Théorie des chaînes de Markov par ex., lorsque échantillonnage Monte Carlo par chaînes de Markov; par ex., en écho à la dynamique de la procédure d'optimisation.

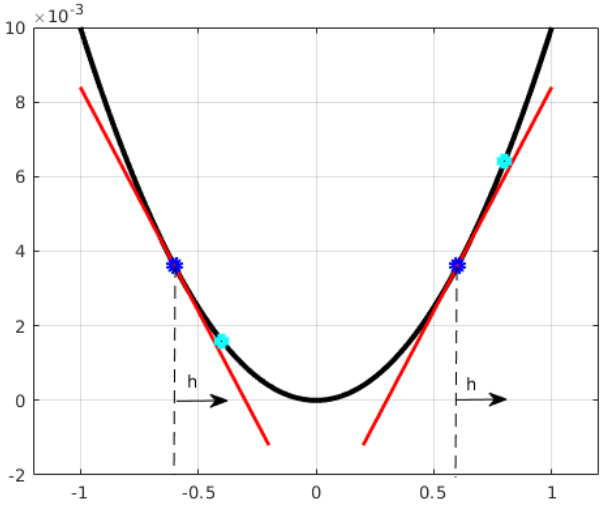
Quelques outils probabilistes
Loi des grands nombres
Autres résultats de Probabilités

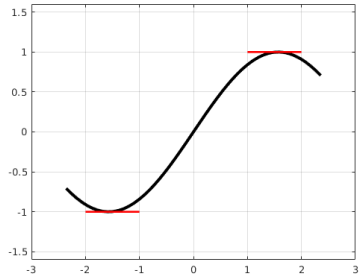
L'aléa pour l'optimisation
Exemples d'algorithmes d'optimisation stochastique
Gradient stochastique
Expectation-Maximisation stochastique
Approximation Stochastique
Le SLAM

L'optimisation pour l'aléatoire : simulation Monte Carlo
Méthodes de Monte Carlo
Taux de reproduction de la Covid19

Conclusion

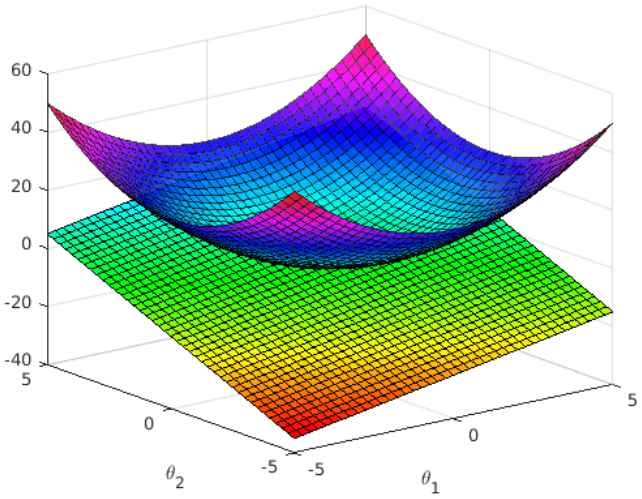
Le gradient





$$F(\theta + h) \approx F(\theta) + h\, F'(\theta)$$

Se déplacer dans la direction $-F'(\theta)$ fait décroître F.



$$\begin{aligned} F(\theta + h) &\approx F(\theta) + h_1\, \partial_{\theta_1} F(\theta) + h_2\, \partial_{\theta_2} F(\theta) \\ &= F(\theta) + \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}^\top \begin{bmatrix} \partial_{\theta_1} F(\theta) \\ \partial_{\theta_2} F(\theta) \end{bmatrix} \\ &= F(\theta) + h^\top \nabla F(\theta) \end{aligned}$$

Se déplacer dans la direction $-\nabla F(\theta)$ fait décroître F.

L'algorithme Gradient stochastique

Robbins-Monro (1951); Kiefer and Wolfowitz (1952); Rosenblatt (1958)

Objectif :

- trouver les zeros d'une dérivée $\theta_\star$ tel que $\nabla F(\theta_\star) = 0$
- dans les situations où $\nabla F(\theta) = \mathbb{E}\left[G(\theta, X)\right]$ non explicite

L'algorithme de gradient

- $\theta_{t+1} = \theta_t - \gamma_{t+1} \nabla F(\theta_t)$ mise en oeuvre impossible

L'idée : une approximation stochastique

- $\nabla F(\theta_t) \approx N^{-1} \sum_{k=1}^{N} G(\theta_t, X_k)$ par tirages de N points $X_1, \cdots, X_N$

L'algorithme de gradient stochastique

Répéter

- tirer (un seul point) X_{t+1} de même loi que X
- $\theta_{t+1} = \theta_t - \gamma_{t+1} G(\theta_t, X_{t+1})$

Ex. Minimisation de risque empirique par Gradient stochastique

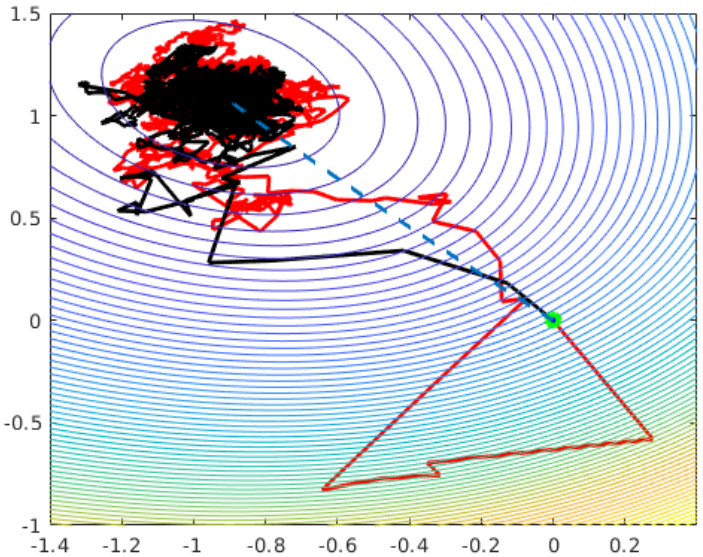
Objectif :

- Minimisation du risque empirique $\quad n^{-1} \sum_{i=1}^{n} \ell(g_\theta(X_i), Y_i)$

L'idée :

- Approximation stochastique du gradient exploitant

$$n^{-1} \sum_{i=1}^{n} \partial_\theta \ell(g_\theta(X_i), Y_i) = \mathbb{E}\left[\partial_\theta \ell(g_\theta(X_I), Y_I) \right] \qquad I \sim \mathcal{U}(\{1, \cdots, n\})$$



L'algorithme Expectation-Maximization (EM)

Dempster et al (1977)

Objectif :

- Minimiser une fonction définie comme l'intégrale d'une fonction positive

$$F(\theta) = -\log \int_{\mathsf{X}} p(x;\theta)\, \nu(\mathrm{d}x)$$

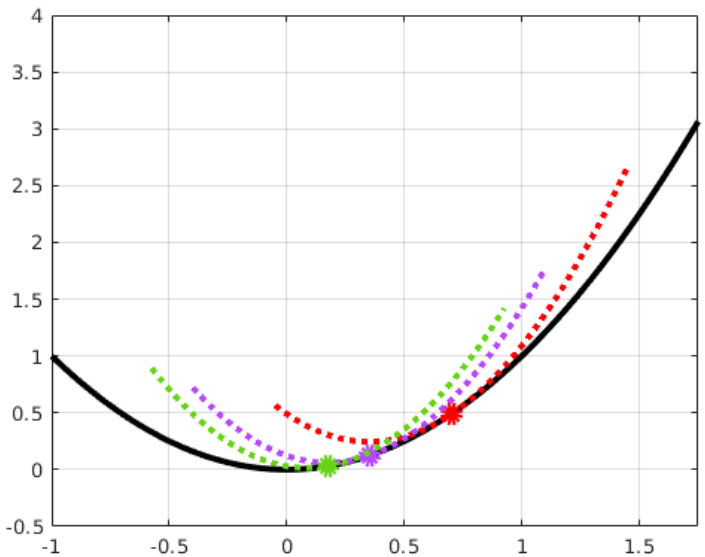
L'idée : répéter

- construire une fct majorante, tangente

$$F(\theta) \leq C_\tau - \int_{\mathsf{X}} \log p(x;\theta)\, \frac{p(x;\tau)}{\int p(z;\tau)\nu(\mathrm{d}z)}\, \nu(\mathrm{d}x)$$

$$= C_\tau - \mathbb{E}_\tau\left[\log p(X;\theta)\right] \qquad X \sim \frac{p(x;\tau)\,\nu(\mathrm{d}x)}{\int p(z;\tau)\,\nu(\mathrm{d}z)}$$

$$C_\tau \text{ t.q. } F(\tau) = C_\tau - \mathbb{E}_\tau\left[\log p(X;\tau)\right]$$



- la minimiser : obtenir $\theta_\star$.

On a toujours

$$F(\theta_\star) \leq F(\tau)$$

↪ algo. Majoration-Minoration

L'algorithme EM stochastique

$$F(\theta) \leq C_\tau - \mathbb{E}_\tau \left[\log p(X;\theta)\right] \qquad X \sim \frac{p(x;\tau)\,\nu(\mathrm{d}x)}{\int p(z;\tau)\,\nu(\mathrm{d}z)}$$

L'algorithme EM Stochastique

- l'EM stochastique remplace la fonction majorante - lorsque non explicite - par une approximation stochastique.

On perd la propriété de décroissance de la fonction objectif $\cdots$

Approximation Stochastique

Robbins and Monro (1951)

Le Gradient stochastique, comme certaines versions de l'EM stochastique, sont des cas particuliers de **l'Approximation Stochastique**

Pour résoudre

$$\theta \in \mathbb{R}^d, \qquad \text{t.q.} \quad \mathbb{E}\left[H(\theta, X)\right] = 0$$

lors l'espérance ne peut pas être calculée de façon exacte.

Choisir une suite de pas d'apprentissage $\{\gamma_{t+1}, t \geq 0\}$
Répéter

- Obtenir un point X_{t+1} de même loi que X
- Poser $\theta_{t+1} = \theta_t + \gamma_{t+1}\, H(\theta_t, X_{t+1})$

Les résultats de convergence sont liés à l'existence d'une fonction de Lyapunov / fonction de descente, pour le champs moyen $\theta \mapsto \mathbb{E}\left[H(\theta, X)\right]$.

Quelques outils probabilistes
Loi des grands nombres
Autres résultats de Probabilités

L'aléa pour l'optimisation
Exemples d'algorithmes d'optimisation stochastique
Gradient stochastique
Expectation-Maximisation stochastique
Approximation Stochastique
Le SLAM

L'optimisation pour l'aléatoire : simulation Monte Carlo
Méthodes de Monte Carlo
Taux de reproduction de la Covid19

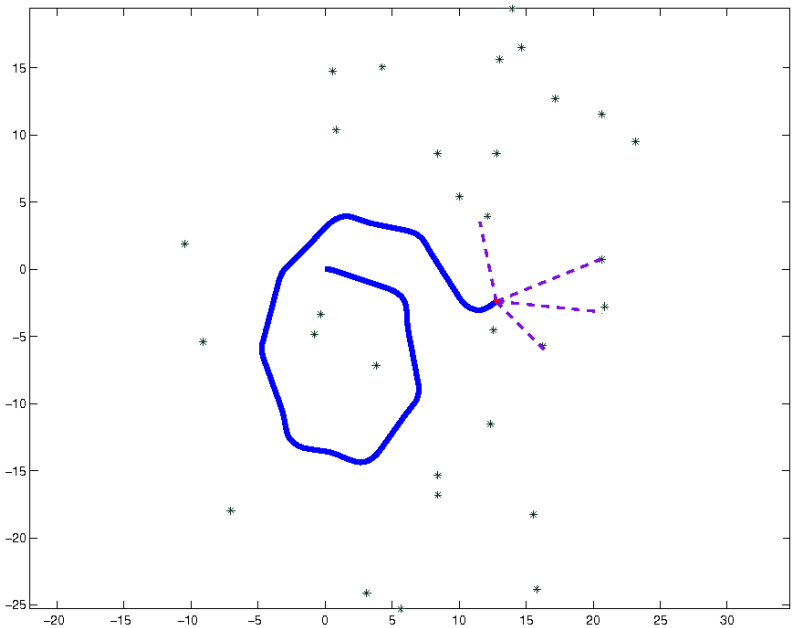
Conclusion

Le SLAM (Simultaneous Localization and Mapping)

Résultats issus des travaux de S. Le Corff (Sorbonne Univ.), G.F. et E. Moulines (Ecole Polytechnique)

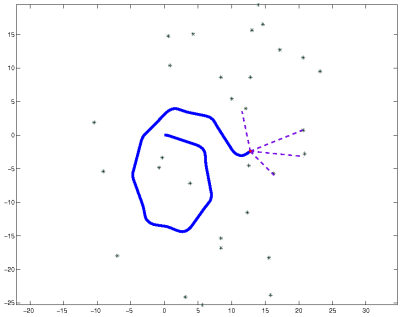
Un robot évolue dans un environnement inconnu, qu'il doit cartographier.

- Cartographie : il apprend la position des objets qu'il détecte.
- Localisation : il estime sa propre position.
- Ces deux tâches sont faites simultanément.



Le modèle

- Un paramètre inconnu
  $\theta = \{\theta_{:,i}, i \leq q\}$: qui collecte les coordonnées cartésiennes des q objets à localiser.



- Un processus aléatoire caché, qui modélise la position du robot à l'instant $\#t$: deux composantes pour les coordonnées (cartésiennes) et une composante pour l'orientation (angulaire)

$$X_t = f(X_{t-1}, U_t)$$

avec contrôles $u_t^\star$ bruités

$$U_t = u_t^\star + \mathcal{N}(0, Q)$$

$$f(x; (v, \psi)) = x + v\, d \begin{bmatrix} \cos(x_3\, \psi) \\ \sin(x_3\, \psi) \\ \sin(\psi)/B \end{bmatrix}$$

- Un processus d'observations, qui modélise ce que le robot *mesure* : pour tout objet $\#i$ observé à l'instant $\#t$

$$Y_{t,i} = h(X_t, \theta_{:,i}) + W_{t,i}$$

$W_{t,i}$ mélange de gaussiennes

$$h(x, \tau) = \begin{bmatrix} \mathrm{dist}(x, \tau) \\ \arctan\left(\frac{\tau_2 - x_2}{\tau_1 - x_1}\right) - x_3 \end{bmatrix}$$

Le critère à optimiser

- Maximisation de la fonction de vraisemblance

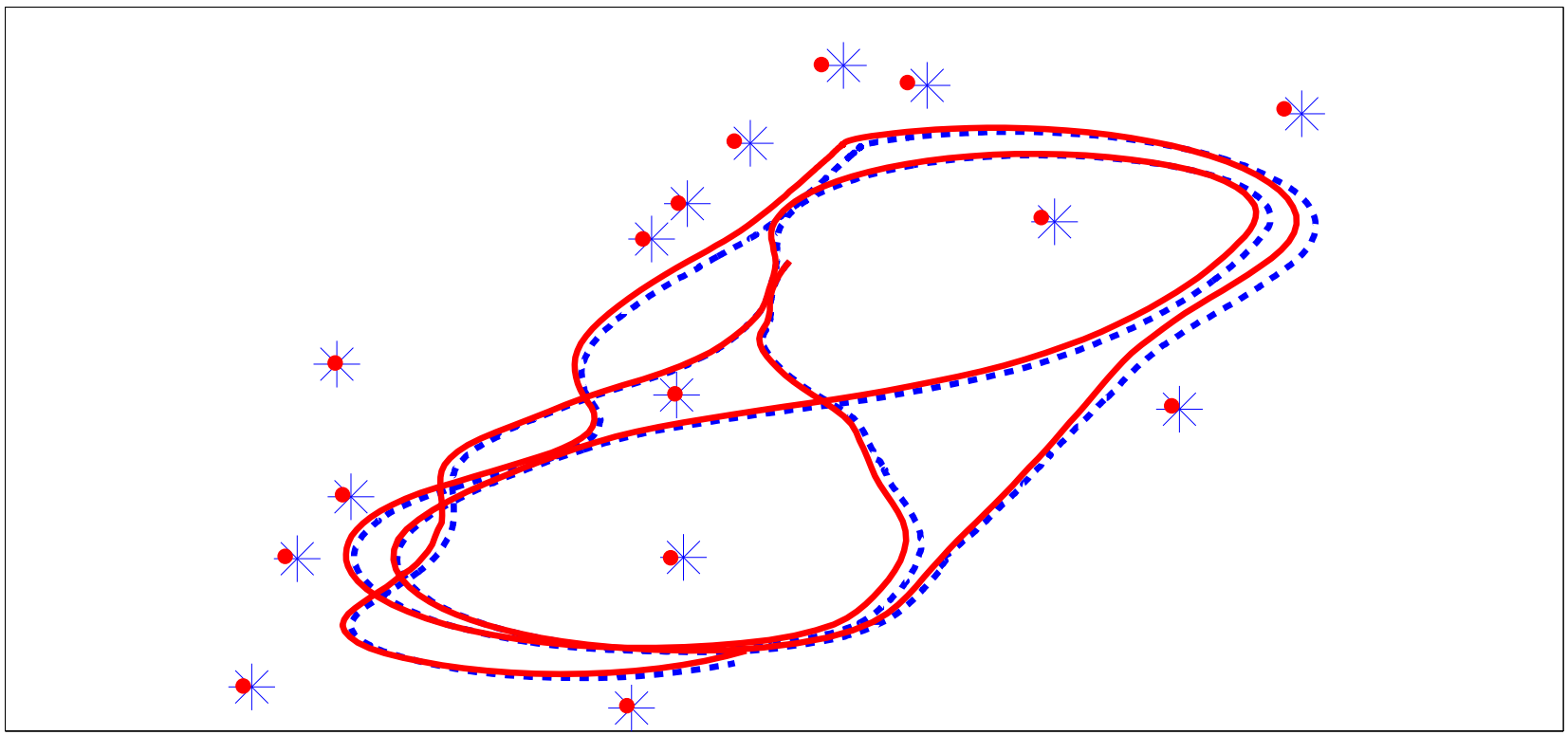
$$\theta \mapsto \ln \text{loi}(Y_1, Y_2, \cdots, Y_T; \theta)$$

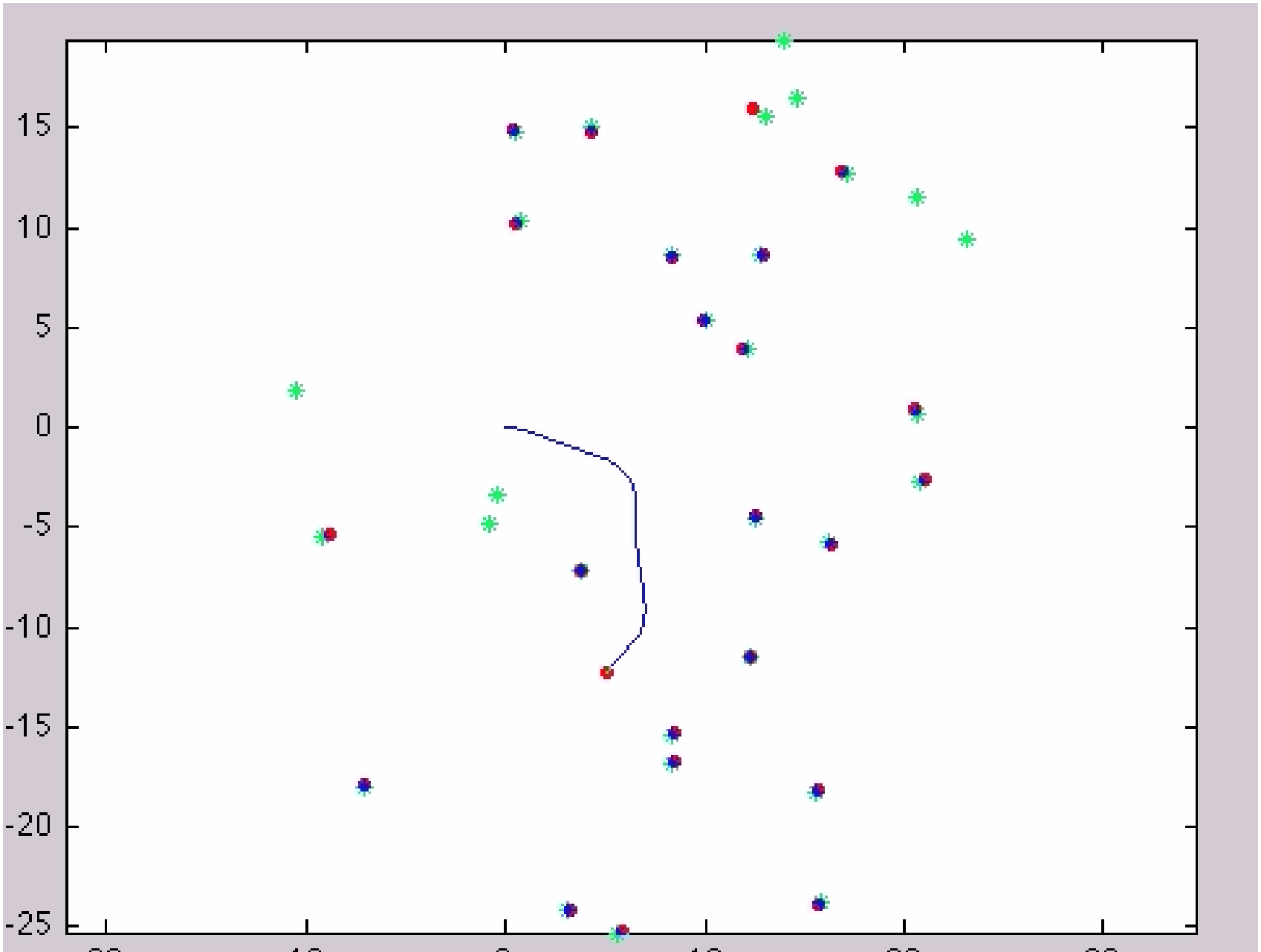
- La fonction objectif est définie comme une intégrale, non explicite

$$\begin{aligned} &\ln \text{loi}(Y_1, Y_2, \cdots, Y_T; \theta) \\ &= \ln \int \text{loi}\,(Y_1, \cdots, Y_T, x_1, \cdots, x_T, z_1, \cdots, z_T; \theta)\ \mathrm{d}x_1 \cdots \mathrm{d}x_T\ \mu(\mathrm{d}z_1) \cdots \mu(\mathrm{d}z_T) \end{aligned}$$

- L'optimisation est réalisée par un algorithme EM **stochastique**.
- L'approximation stochastique nécessite de simuler le processus caché $(X_1, \cdots, X_T, Z_1, \cdots, Z_T)$ sous une loi dépendant des observations $Y_1, \cdots, Y_T$: **filtrage particulaire**.
- Gestion de l'**apprentissage en ligne** Block Online EM (BOEM)

Résultats




15
10
5
0
-5
-10
-15
-20
-25

Quelques outils probabilistes
Loi des grands nombres
Autres résultats de Probabilités

L'aléa pour l'optimisation
Exemples d'algorithmes d'optimisation stochastique
Gradient stochastique
Expectation-Maximisation stochastique
Approximation Stochastique
Le SLAM

L'optimisation pour l'aléatoire : simulation Monte Carlo
Méthodes de Monte Carlo
Taux de reproduction de la Covid19

Conclusion

"Monte Carlo"

- Comte de Buffon (18ème) : l'aiguille de Buffon, pour l'estimation de π.
- E. Fermi (env. 1930) : physicien, exploitation de l'énergie nucléaire; prix Nobel de Physique en 1938.
- S. Ulam (fin 1940) : "The first thoughts and attempts I made to practice [the Monte Carlo Method] were suggested by a question which occurred to me in 1946 as I was convalescing from an illness and playing solitaires. The question was what are the chances that a Canfield solitaire laid out with 52 cards will come out successfully? After spending a lot of time trying to estimate them by pure combinatorial calculations, I wondered whether a more practical method than "abstract thinking" might not be to lay it out say one hundred times and simply observe and count the number of successful plays. This was already possible to envisage with the beginning of the new era of fast computers, and I immediately thought of problems of neutron diffusion and other questions of mathematical physics, and more generally how to change processes described by certain differential equations into an equivalent form interpretable as a succession of random operations. Later [in 1946], I described the idea to John von Neumann, and we began to plan actual calculations." Travaux secrets, nom de code "Monte Carlo", suggéré par N. Metropolis.

Méthodes de Monte Carlo

Une loi de probabilité ν sur un espace X

- un objet, fonction définie sur une tribu, famille de sous-ensembles de X t.q.

$$\nu(A) \in [0,1]\,,$$
$$\nu(\mathsf{X}) = 1,$$
$$\nu(\cup_{i\geq 1} A_i) = \sum_{i\geq 1} \nu(A_i) \quad \text{pour } A_i, A_j \text{ disjoints}$$

Les méthodes de Monte Carlo : **étant donnée une loi ν**

- exploitent la LGN

$$\int f(x)\,\nu(\mathrm{d}x) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} f(X_k)$$

pour approcher ν par un nuage de points $\{X_1, X_2, \cdots, X_N\}$.

- sont un outil numérique pour comprendre la loi ν, et approcher des quantités non calculables explicitement relatives à ν (espérance, quantiles, probabilités d'événements, $\cdots$).

Définition de la loi ν

- expression mathématique explicite

  ex. $p(x)\,\mu(\mathrm{d}x)$ pour $x \in \mathsf{X} \subseteq \mathbb{R}^d$.

- expression mathématique explicite, à une constante de normalisation près

  ex. $p(x)\,\mu(\mathrm{d}x)$ pour $x \in \mathsf{X} \subseteq \mathbb{R}^d$ mais $\int_{\mathsf{X}} p(x)\mu(\mathrm{d}x) \neq 1$

- définition non explicite. Par ex., une description du mécanisme ayant mené à la réalisation du phénomène aléatoire observée

Echantillonnage Monte Carlo exact

Certaines lois sont *standard* : il existe des algorithmes pour obtenir des points $x_1, x_2, \cdots$ vus comme la réalisation de $X_1, X_2, \cdots$ t.q. $X_k \sim \nu$.

A partir d'un générateur rand de loi uniforme sur $[0,1]$

loi uniforme sur $\{1, \cdots, 6\}$

$$U \leftarrow \lceil 6\ \mathrm{rand} \rceil$$

loi exponentielle, de paramètre λ

$$X \leftarrow -\ln(1-\mathrm{rand})/\lambda$$

loi de type mélange

si $\mathrm{rand} \leq \omega$,
$X \leftarrow \mathrm{blackbox}$
sinon $X \leftarrow 0$

loi gaussienne $\mathcal{N}_2(0, \mathrm{I})$

Répéter
$X \leftarrow 2\,\mathrm{rand} - 1$
$Y \leftarrow 2\,\mathrm{rand} - 1$
Tant que $X^2 + Y^2 > 1$
$S \leftarrow X^2 + Y^2$
$Z \leftarrow \sqrt{-2\ln(S)/S}$
$(G_1, G_2) \leftarrow (ZX, ZY)$

Echantillonnage Monte Carlo par Chaînes de Markov (MCMC)

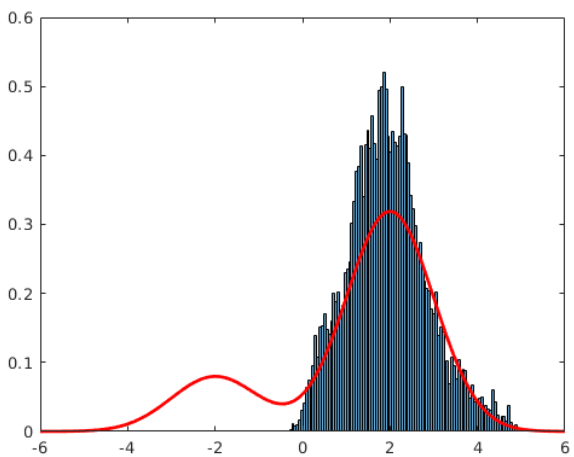
- La loi des grands nombres, entre autres résultats de probabilité, reste vraie lorsque $\{X_1, X_2, \cdots, \}$ est une Chaîne de Markov ayant de "bonnes" propriétés.

$$\frac{1}{N}\sum_{k=1}^{N} f(X_k) \to \int f(x)\nu(\mathrm{d}x).$$

- Chaîne de Markov (ChM) :

$$\text{loi}(X_t|X_1, \cdots, X_{t-1}) = \text{loi}(X_t|X_{t-1})$$

- ChM ergodique : la loi de X_t converge vers une loi $\nu(\mathrm{d}x)$, invariante pour le mécanisme de transition.



Hastings-Metropolis

Algorithme de Hastings-Metropolis, pour approcher $p(x)\mu(\mathrm{d}x)$
Choisir une loi de proposition $q(x,y)\mu(\mathrm{d}y)$

Répéter

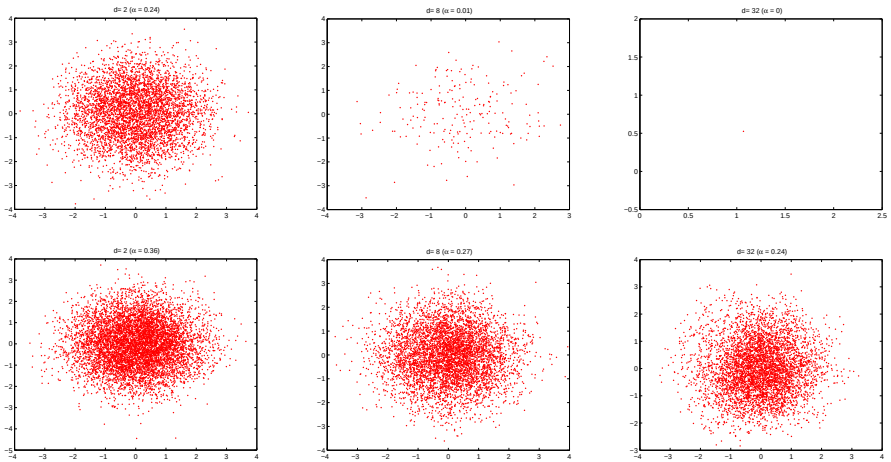
- Proposer $Y \sim q(X_t, y)\mu(\mathrm{d}y)$
- Calculer la probabilité d'accepter ce candidat

$$\alpha(X_t, Y) = 1 \wedge \left(\frac{p(Y)}{p(X_t)} \frac{q(Y, X_t)}{q(X_t, Y)} \right)$$

- Accepter ou pas ce candidat :
  $X_{t+1} = Y$ avec probabilité $\alpha(X_t, Y)$
  $X_{t+1} = X_t$ sinon.

Hastings-Metropolis à marche aléatoire symétrique

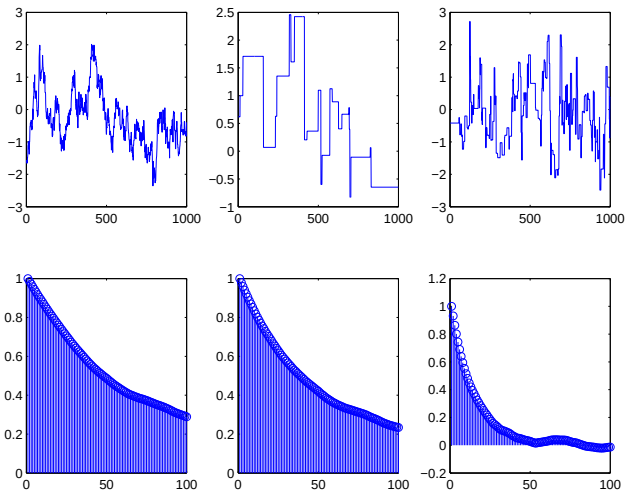
$$q(x,y) = \tilde{q}(|x-y|) \qquad \text{ex.} \quad y = x + \mathcal{N}(0,C)$$



Composante 1 d'une chaîne $X_1, X_2, \cdots$, "visant" $\mathcal{N}_d(0,I)$ pour $d \in \{2,8,32\}$: le candidat est $Y = X_t + \mathcal{N}_d(0,\sigma^2 I)$.

σ ne dépend pas de d (haut) et σ est de la forme $c/\sqrt{d}$ (bas)

- Résultats théoriques :
  $\alpha_\star = 0.234 \qquad C = 2.38^2\, C_\pi/d$
- Critère à optimiser au cours de l'algorithme pour changer ses paramètres d'implémentation : échantillonner, optimiser, échantillonner, $\cdots$



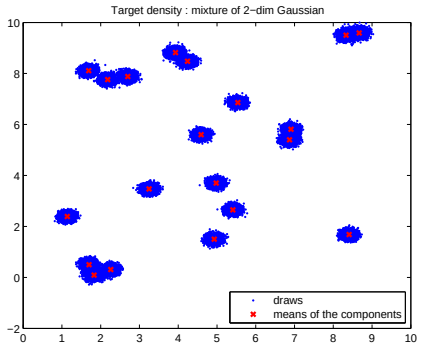
Optimisation pour améliorer l'algorithme MCMC

MCMC adaptatif et/ou en interaction
Echantillonner pour approcher la loi ν, à l'aide d'une famille de transitions markoviennes P_θ.
Résoudre : $\theta_\star$ t.q. $\int H(x,\theta)\nu(\mathrm{d}x) = 0$

Répéter

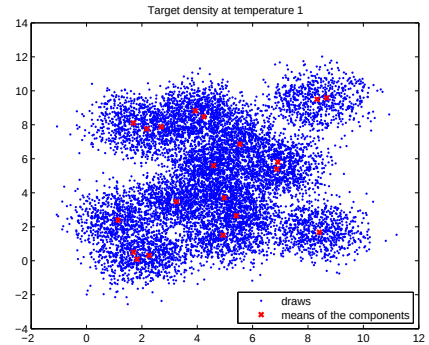
- une itération d'un algorithme d'optimisation stochastique cherchant le zero, et exploitant X_t : obtenir θ_t
- une itération d'un échantillonneur exploitant θ_t : obtenir $X_{t+1} \sim P_{\theta_t}(X_t, \cdot)$

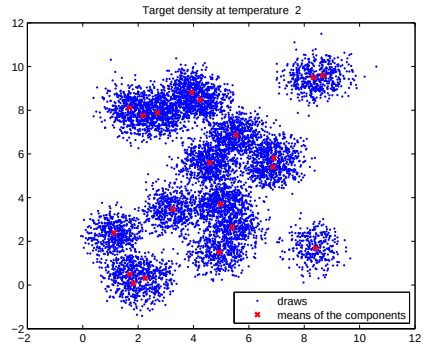
Exemple 1, "Adaptive Equi-Energy"

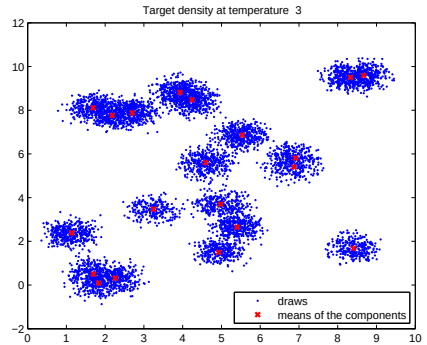


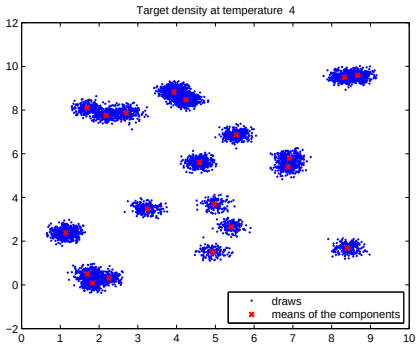
Apprendre des quantiles : $\int 1_{(0,\theta_\star]}(x)\,\nu(\mathrm{d}x) - \alpha = 0$

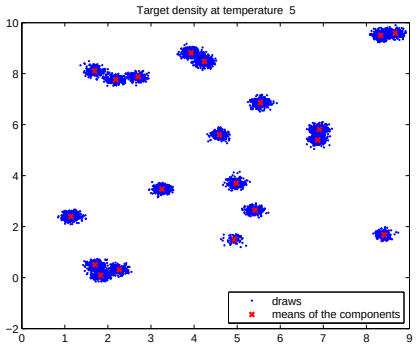
- (gauche) La loi à apprendre $p(x)\,\mathrm{d}x$
- (dessous) Quatre processus auxiliaires, approchant des lois tempérées $p^{1/T}(x)$ à des températures de moins en moins chaudes; l'approximation de ν; un simple HM.

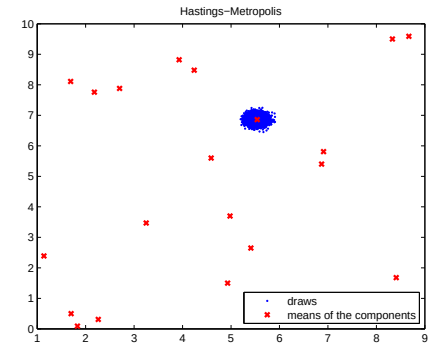




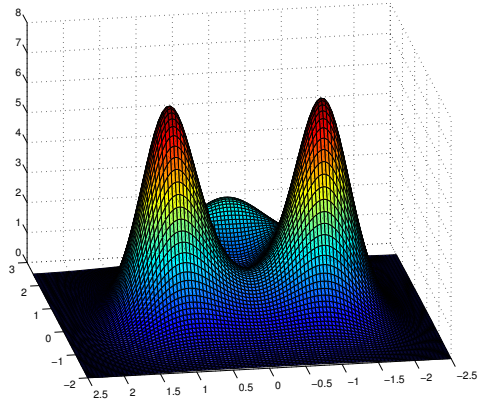


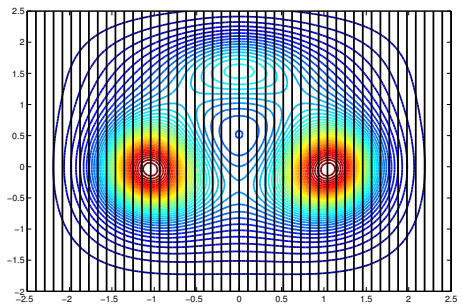






Exemple 2, "Adaptive Biasing Force" de type Wang-Landau





- Estimer les poids de chaque strate $\#i$

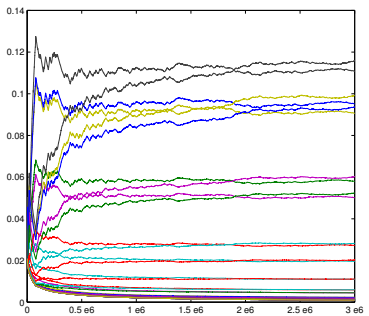
$$\theta_\star(i) = \int 1_{\mathsf{X}_i}(x)\nu(\mathrm{d}x)$$

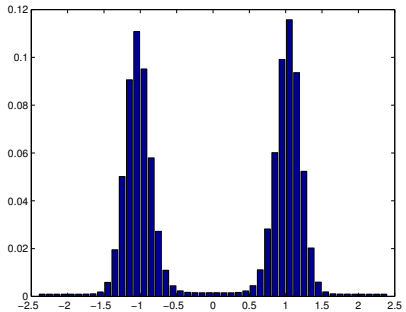
- A l'aide d'une chaîne qui vise une loi $\nu_{\theta_\star}$ obtenue en déformant ν; la déformation indexée par $\theta_\star$ donne le même poids à toutes les strates.

$$\forall i,\ \sum_{k=1}^{I} \int_{\mathsf{X}_k} \frac{\theta(i)}{\theta(k)} \left(1_{\mathsf{X}_i}(x) - \sum_{j=1}^{J} \theta(j) 1_{\mathsf{X}_j}(x) \right) \nu(\mathrm{d}x) = 0$$

(haut) densité; strates

(droite) evolution de l'apprentissage de ces poids; les poids appris





Quelques outils probabilistes
Loi des grands nombres
Autres résultats de Probabilités

L'aléa pour l'optimisation
Exemples d'algorithmes d'optimisation stochastique
Gradient stochastique
Expectation-Maximisation stochastique
Approximation Stochastique
Le SLAM

L'optimisation pour l'aléatoire : simulation Monte Carlo
Méthodes de Monte Carlo
Taux de reproduction de la Covid19

Conclusion

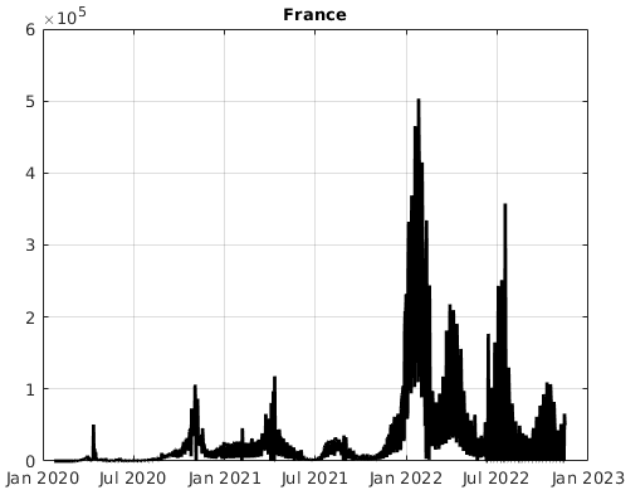
Taux de reproduction d'une épidémie

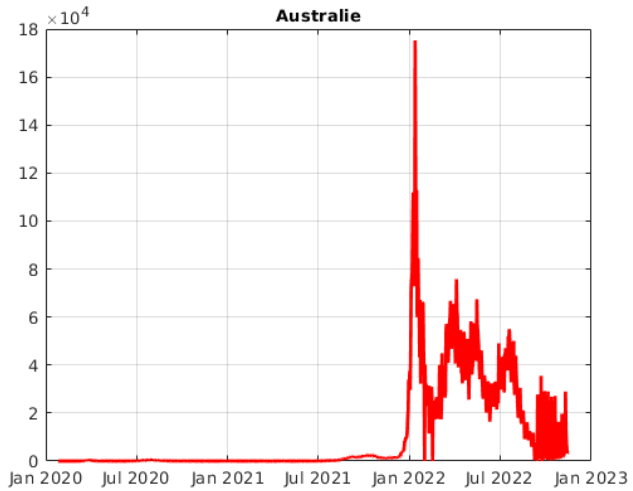
Résultats issus des travaux de P. Abry (CNRS, Lyon), G.F., B. Pascal (CNRS, Nantes) et N. Pustelnik (CNRS, Lyon)

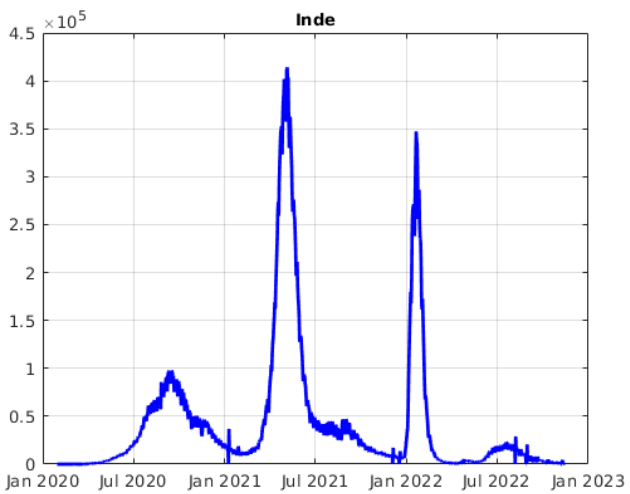
- La récente épidémie de Covid19 et les crises engendrées (santé, société, économie) montrent la nécessité de surveiller l'évolution d'une pandémie.
- Menée en contexte intrapandémie, cette surveillance aide aussi les pouvoirs publics dans la prise de décisions sanitaires et donc politiques (ex. vaccination, restrictions de déplacements, confinements).
- Les données intrapandémie ("temps réel") sont souvent de qualité limitée. Pour Covid19 : comptes aberrants parfois négatifs, comptes non reportés, comptes reportés les jours suivants, $\cdots$
- Modèle classique : modèles compartimentaux (sains, infectés, guéris, décédés). Ils souffrent de la complexité du modèle (nombre de compartiments et donc des paramètres décrivant le modèle) et de la difficulté de les calibrer – difficulté accrue par la faible qualité des données.
- Objectif : modèle adapté à la surveillance intrapandémie.
- Approche : suivre le **taux de reproduction** = nombre moyen d'individus qui seront infectés par une personne déjà infectée.

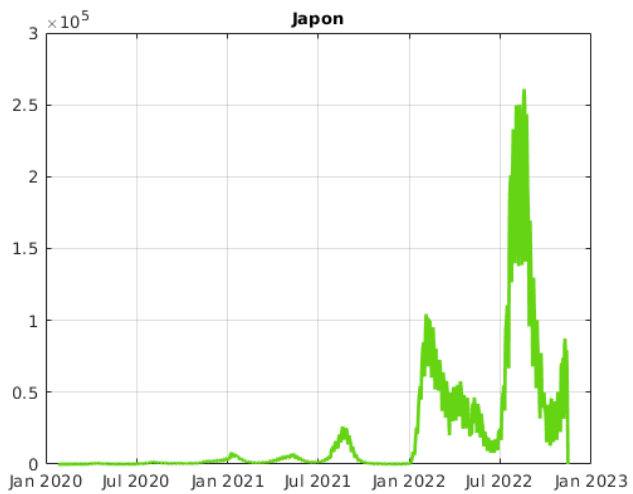
Les données : nombre de nouveaux infectés, par jour, par pays

Données disponibles sur le site de Johns Hopkins University (USA)



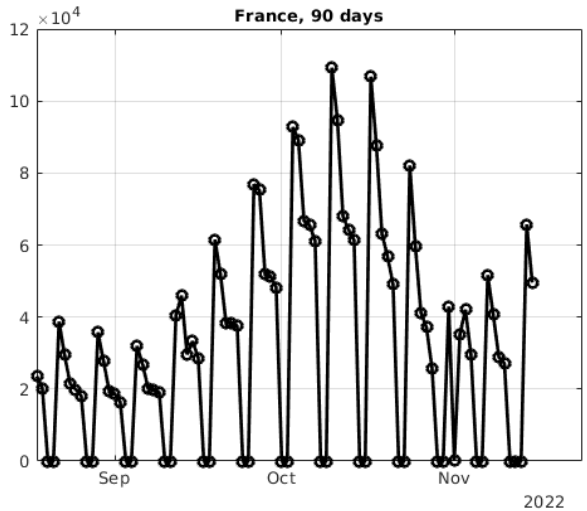


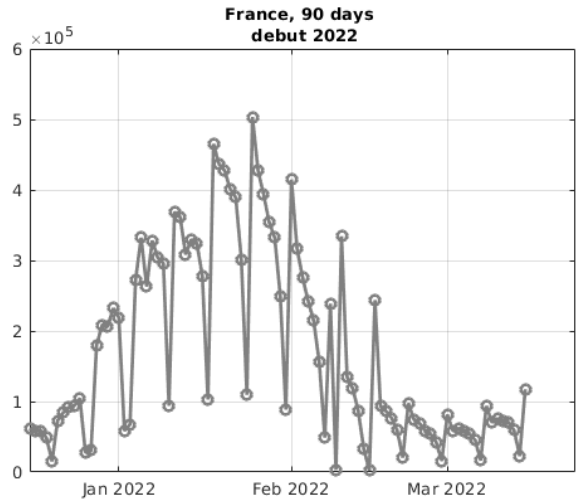


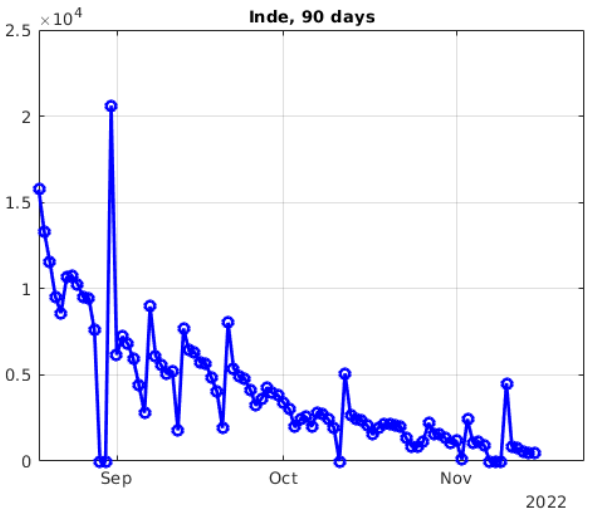


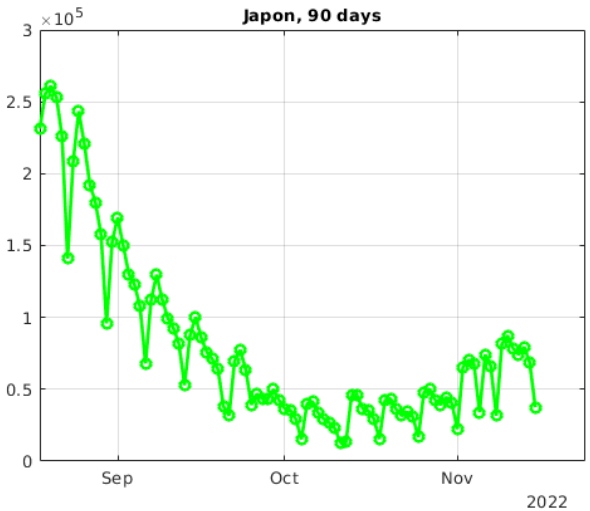
Les données : zoom sur 90 jours

Données disponibles sur le site de Johns Hopkins University (USA)







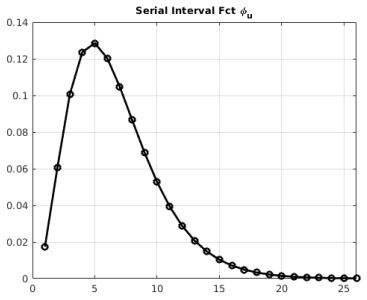


Le modèle

- Les comptes :

$$\text{loi}(Z_t|Z_1,\cdots,Z_{t-1}) = \mathcal{P}\left(\mathsf{R}_t\Phi_t + \mathsf{O}_t\right)$$

$$\Phi_t = \sum_{u=1}^{\tau} \phi(u)\mathsf{Z}_{t-u}$$



- Le taux de reproduction R_t : linéaire par morceaux

$$|\mathsf{R}_t - 2\mathsf{R}_{t-1} + \mathsf{R}_{t-2}| \qquad \text{parcimonieux} \qquad \mathsf{R}_t > 0$$

- Des erreurs (sous/sur-évaluations) :

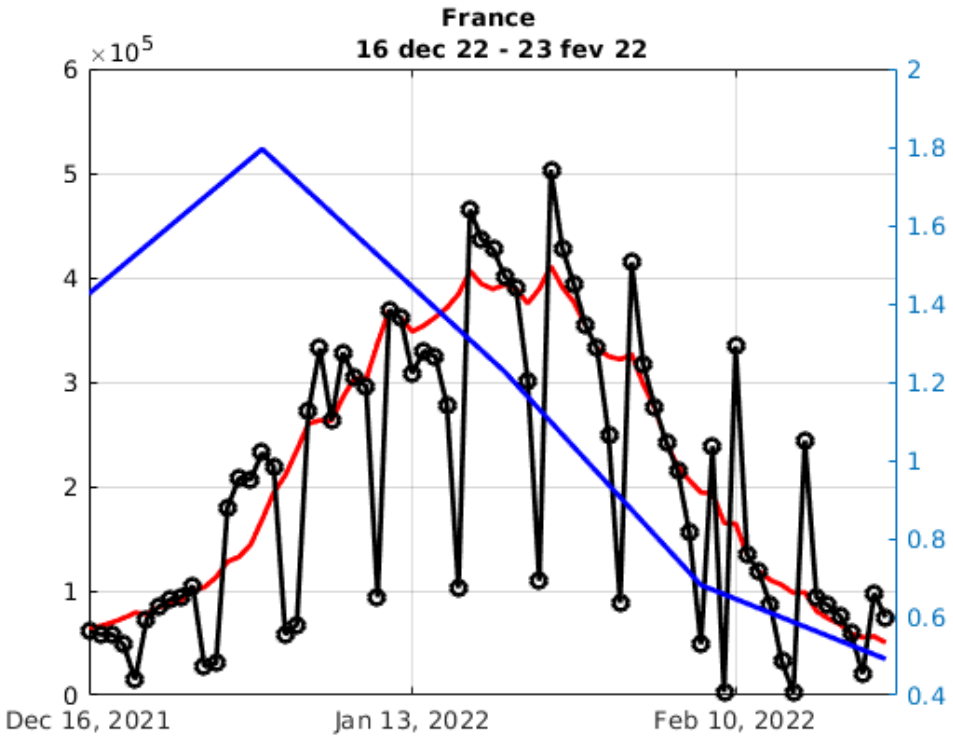
$$|\mathsf{O}_t| \qquad \text{parcimonieux} \qquad \mathsf{O}_t \in \mathbb{R}$$

Une **loi de probabilité pour** $(\mathsf{R}_1,\cdots,\mathsf{R}_T,\mathsf{O}_1,\cdots,\mathsf{O}_T)$

$$\exp\left(\sum_{t=1}^{T}\left(\mathsf{Z}_t\ln(\mathsf{R}_t\Phi_t+\mathsf{O}_t)-\mathsf{R}_t\Phi_t+\mathsf{O}_t\right)-\lambda_{\mathsf{R}}\sum_{t=1}^{T-2}|\mathsf{R}_t-2\mathsf{R}_{t+1}+\mathsf{R}_{t+2}|-\lambda_0\sum_{t=1}^{T}|\mathsf{O}_t|\right)$$

Optimisation

- Le mode de la loi est calculable par un algorithme d'optimisation convexe, primal-dual.



Proximal et Gradient pour un MCMC efficace

- Loi avec un seul mode, log-concave
- Exploration de la loi pour calcul d'intervalles de crédibilité à 95% – à l'aide des quantiles de la loi, appris par MCMC.

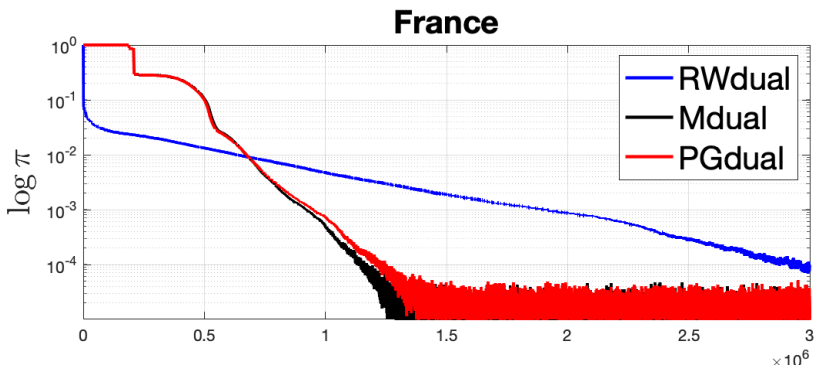
Améliorer Hastings-Metropolis à marche aléatoire symétrique :

- Si la densité p est régulière

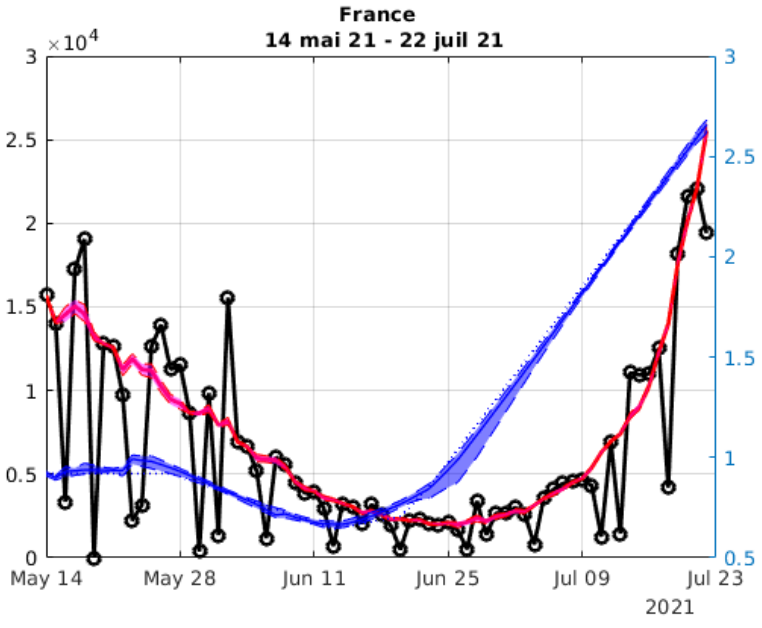
$$Y = X_t + \gamma_{t+1} \nabla \ln p(X_t) + \mathcal{N}(0, 2\gamma_{t+1} \mathrm{I}) \qquad \text{Langevin Monte Carlo}$$

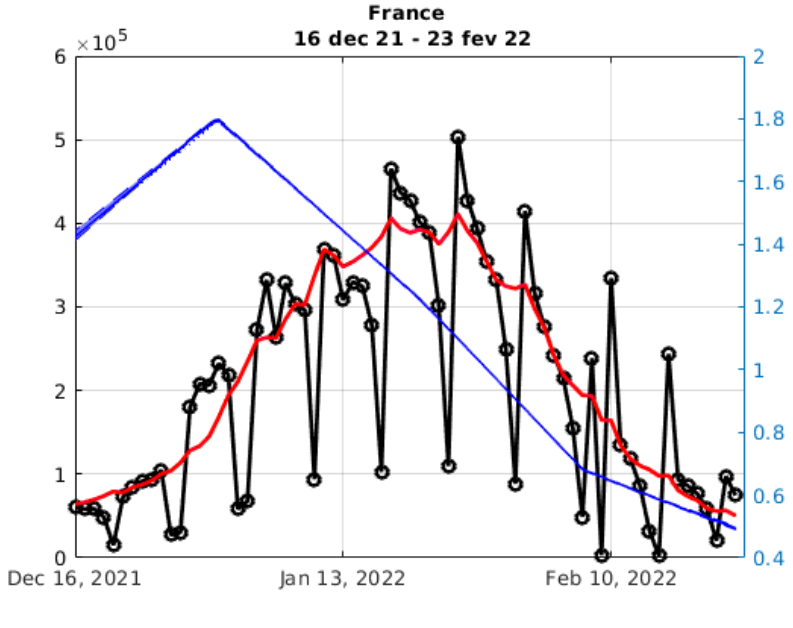
- Sinon, en exploitant la concavité de la partie non régulière

$$Y = \text{Prox, Grad}(X_t, p(X_t)) + \mathcal{N} \qquad \text{Proximal-Langevin Monte Carlo}$$



Résultats


France
14 mai 21 - 22 juil 21
×10^4
3
2.5
2
1.5
1
0.5
0
May 14
May 28
Jun 11
Jun 25
Jul 09
Jul 23
2021


France
16 dec 21 - 23 fev 22
×10^5
6
5
4
3
2
1
0
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
Dec 16, 2021
Jan 13, 2022
Feb 10, 2022

Quelques outils probabilistes
- Loi des grands nombres
- Autres résultats de Probabilités

L'aléa pour l'optimisation
- Exemples d'algorithmes d'optimisation stochastique
  - Gradient stochastique
  - Expectation-Maximisation stochastique
  - Approximation Stochastique
- Le SLAM

L'optimisation pour l'aléatoire : simulation Monte Carlo
- Méthodes de Monte Carlo
- Taux de reproduction de la Covid19

Conclusion

Conclusion

- Illustration de la pertinence d'associer "optimisation" et "aléatoire"
- Optimisation stochastique
- Méthodes de Monte Carlo efficaces
- Les méthodes d'échantillonnage permettent de réduire le coût computationnel de procédures d'optimisation, et/ou de rendre possible leur mise en oeuvre : en proposant une approximation aléatoire de la procédure d'optimisation déterministe.
- Les méthodes d'optimisation permettent d'améliorer les procédures d'échantillonnage Monte Carlo : pour guider l'exploration aléatoire vers les zones de forte probabilité, et/ou pour apprendre au cours de l'algorithme comment améliorer l'efficacité de l'algorithme.
- Dans les deux cas, les algorithmes classiques sont "perturbés" (approximations d'étapes d'optimisation; évolution des mécanismes de transition dans la procédure Monte Carlo).
- Il est nécessaire d'établir le bien-fondé de ces algorithmes stochastiques (convergence, vitesse, quantification des erreurs, etc).