Irreducibility and convergence of nonlinear state-space models

Application: CMA-ES

Armand Gissler

Tuesday 3rd October, 2023

CMAP, École polytechnique & Inria
(Advisors: Anne Auger & Nikolaus Hansen)

Control model [Meyn & Caines 1991, Meyn & Tweedie 2012]

$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad \textbf{(CM(F))}$$

Control model [Meyn & Caines 1991, Meyn & Tweedie 2012]

$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad (\mathbf{CM}(\mathbf{F}))$$

- $F\colon \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ **smooth** ($\mathcal{C}^\infty$)

Control model [Meyn & Caines 1991, Meyn & Tweedie 2012]

$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad \textbf{(CM(F))}$$

- $F\colon \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ **smooth** ($\mathcal{C}^\infty$)
- $\{v_{k+1}\}$ i.i.d. with a l.s.c. density p

Control model [Meyn & Caines 1991, Meyn & Tweedie 2012]

$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad \textbf{(CM(F))}$$

- $F\colon \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ **smooth** ($\mathcal{C}^\infty$)
- $\{v_{k+1}\}$ i.i.d. with a l.s.c. density p
- $\mathsf{X} \subset \mathbb{R}^n$, $\mathsf{V} \subset \mathbb{R}^p$ **open**.

Control model [Meyn & Caines 1991, Meyn & Tweedie 2012]

$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad \textbf{(CM(F))}$$

- $F\colon \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ **smooth** ($\mathcal{C}^\infty$)
- $\{v_{k+1}\}$ i.i.d. with a l.s.c. density p
- $\mathsf{X} \subset \mathbb{R}^n$, $\mathsf{V} \subset \mathbb{R}^p$ **open**.

Theorem

Control model [Meyn & Caines 1991, Meyn & Tweedie 2012]

$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad \textbf{(CM(F))}$$

- $F \colon \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ **smooth** ($\mathcal{C}^\infty$)
- $\{v_{k+1}\}$ i.i.d. with a l.s.c. density p
- $\mathsf{X} \subset \mathbb{R}^n$, $\mathsf{V} \subset \mathbb{R}^p$ **open**.

Theorem

(i) *If* **CM(F)** *is* **forward accessible**, *then* $\{\theta_k\}_{k\in\mathbb{N}}$ *is a T-chain ;*

Control model [Meyn & Caines 1991, Meyn & Tweedie 2012]

$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad (\mathbf{CM}(\mathbf{F}))$$

- $F\colon \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ **smooth** ($\mathcal{C}^\infty$)
- $\{v_{k+1}\}$ i.i.d. with a l.s.c. density p
- $\mathsf{X} \subset \mathbb{R}^n$, $\mathsf{V} \subset \mathbb{R}^p$ **open**.

Theorem

(i) *If* **CM(F)** *is* **forward accessible**, *then* $\{\theta_k\}_{k\in\mathbb{N}}$ *is a T-chain ;*

(ii) *then,* $\{\theta_k\}_{k\in\mathbb{N}}$ *is irreducible* $\Leftrightarrow$ *there exists a* **globally attracting state** *;*

Control model [Meyn & Caines 1991, Meyn & Tweedie 2012]

$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad (\mathbf{CM(F)})$$

- $F : \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ **smooth** ($\mathcal{C}^\infty$)
- $\{v_{k+1}\}$ i.i.d. with a l.s.c. density p
- $\mathsf{X} \subset \mathbb{R}^n$, $\mathsf{V} \subset \mathbb{R}^p$ **open**.

Theorem

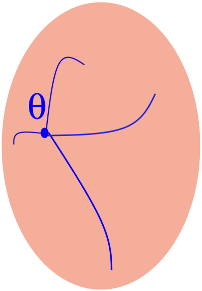
(i) *If* **CM(F)** *is* **forward accessible**, *then* $\{\theta_k\}_{k\in\mathbb{N}}$ *is a T-chain ;*

(ii) *then,* $\{\theta_k\}_{k\in\mathbb{N}}$ *is irreducible* $\Leftrightarrow$ *there exists a* **globally attracting state** *;*

(iii) *the (a)periodicity of* **CM(F)** *is equivalent to the (a)periodicity of* $\{\theta_k\}_{k\in\mathbb{N}}$.

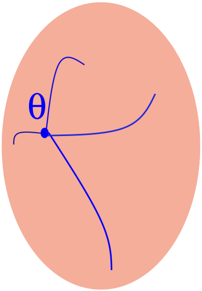
Forward accessibility

$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad \textbf{(CM(F))}$$



Forward accessibility

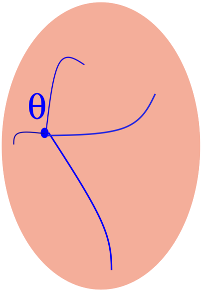
$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad \textbf{(CM(F))}$$
$$= F_{k+1}(\theta_0, v_1, \ldots, v_{k+1})$$



Forward accessibility

$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad \textbf{(CM(F))}$$
$$= F_{k+1}(\theta_0, v_1, \ldots, v_{k+1})$$

for $v_1, \ldots, v_{k+1} \in \mathcal{O} = \mathrm{supp}\ p$.



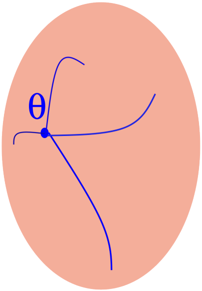
Forward accessibility

$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad \textbf{(CM(F))}$$
$$= F_{k+1}(\theta_0, v_1, \ldots, v_{k+1})$$

for $v_1, \ldots, v_{k+1} \in \mathcal{O} = \mathrm{supp}\ p$.

$A_+(\theta) = \{F_k(\theta, v_1, \ldots, v_k) \mid k \in \mathbb{N},\ v_1, \ldots, v_k \in \mathcal{O}\}$

CM(F) is *forward accessible* when $A_+(\theta)$ has nonempty interior



Forward accessibility

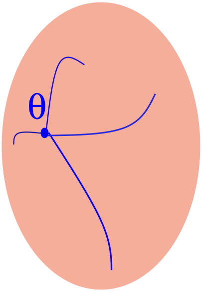
$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad \textbf{(CM(F))}$$
$$= F_{k+1}(\theta_0, v_1, \ldots, v_{k+1})$$

for $v_1, \ldots, v_{k+1} \in \mathcal{O} = \mathrm{supp}\ p$.

$A_+(\theta) = \{F_k(\theta, v_1, \ldots, v_k) \mid k \in \mathbb{N},\ v_1, \ldots, v_k \in \mathcal{O}\}$

CM(F) is *forward accessible* when $A_+(\theta)$ has nonempty interior

$$C_k = \mathrm{Jac}_{v_{1..k}} F_k(\theta, v_1, \ldots, v_k)$$

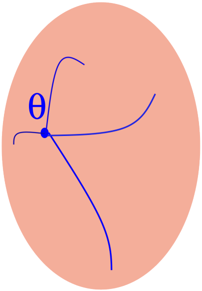


Forward accessibility

$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad \textbf{(CM(F))}$$
$$= F_{k+1}(\theta_0, v_1, \ldots, v_{k+1})$$

for $v_1, \ldots, v_{k+1} \in \mathcal{O} = \mathrm{supp}\ p$.

$A_+(\theta) = \{F_k(\theta, v_1, \ldots, v_k) \mid k \in \mathbb{N},\ v_1, \ldots, v_k \in \mathcal{O}\}$



CM(F) is *forward accessible* when $A_+(\theta)$ has nonempty interior

$$C_k = \mathrm{Jac}_{v_{1..k}} F_k(\theta, v_1, \ldots, v_k)$$

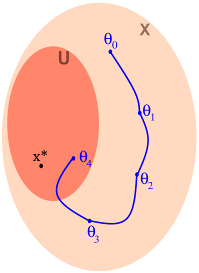
Then:
CM(F) is forward accessible
$\Leftrightarrow \forall\theta,\ \exists v_1, \ldots, v_k \in \mathcal{O},\ \mathrm{rank}\ C_k = n.$

Globally attracting state

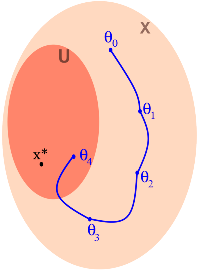
x^* is globally attracting when

$\forall \theta_0, \exists v_1, \dots, v_k \in \mathcal{O} = \mathrm{supp}\ p$,

$$F_k(\theta_0, v_1, \dots, v_k) \in \mathrm{Neighborhood}(x^*)$$



Globally attracting state



x^* is globally attracting when
$\forall \theta_0, \exists v_1, \ldots, v_k \in \mathcal{O} = \mathrm{supp}\ p,$

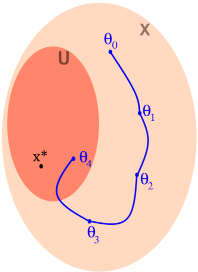
$$F_k(\theta_0, v_1, \ldots, v_k) \in \mathrm{Neighborhood}(x^*)$$

If

$$C_k(x^*, v_1, \ldots, v_k) = \mathrm{Jac}_{v_{1..k}} F_k(x^*, v_1, \ldots, v_k)$$

is of rank n

Globally attracting state



x^* is globally attracting when

$\forall \theta_0, \exists v_1, \ldots, v_k \in \mathcal{O} = \mathrm{supp}\ p,$

$$F_k(\theta_0, v_1, \ldots, v_k) \in \mathrm{Neighborhood}(x^*)$$

If

$$C_k(x^*, v_1, \ldots, v_k) = \mathrm{Jac}_{v_{1..k}} F_k(x^*, v_1, \ldots, v_k)$$

is of rank n,

then $\forall \theta$, $\exists u_1, \ldots, u_j \in \mathcal{O}$,

$$\mathrm{rank}\ C_k(\theta, u_1, \ldots, u_j) = n$$

Aperiodicity of $(\mathbf{CM}(\mathbf{F}))$

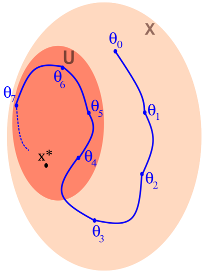
$(\mathbf{CM}(\mathbf{F}))$ is aperiodic $\Leftrightarrow \exists$ a steadily attracting state

Aperiodicity of (CM(F))

(**CM**(**F**)) is aperiodic ⇔ ∃ a steadily attracting state

θ^* is steadily attracting when for any $\theta_0 \in \mathsf{X}$,
$\exists v_1, v_2, \cdots \in \mathcal{O} = \mathrm{supp}\ p$,

$$\lim_{k \to \infty} F_k(\theta_0, v_1, \ldots, v_k) = \theta^*$$



Control model [Meyn & Caines 1991, Meyn & Tweedie 2012]

$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad (\mathbf{CM}(\mathbf{F}))$$

- $F\colon \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ **smooth** ($\mathcal{C}^\infty$)
- $\{v_{k+1}\}$ i.i.d. with a l.s.c. density p
- $\mathsf{X} \subset \mathbb{R}^n$, $\mathsf{V} \subset \mathbb{R}^p$ **open**.

Theorem

(i) *If* **CM(F)** *is* **forward accessible**, *then* $\{\theta_k\}_{k\in\mathbb{N}}$ *is a T-chain ;*

(ii) *then,* $\{\theta_k\}_{k\in\mathbb{N}}$ *is irreducible* $\Leftrightarrow$ *there exists a* **globally attracting state** *;*

(iii) *the (a)periodicity of* **CM(F)** *is equivalent to the (a)periodicity of* $\{\theta_k\}_{k\in\mathbb{N}}$.

Control model [Meyn & Caines 1991, Meyn & Tweedie 2012]

$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad \textbf{(CM(F))}$$

- $F : \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ **smooth** ($\mathcal{C}^\infty$)
- $\{v_{k+1}\}$ i.i.d. with a l.s.c. density p
- $\mathsf{X} \subset \mathbb{R}^n$, $\mathsf{V} \subset \mathbb{R}^p$ **open**.

Theorem
Suppose that θ^* is steadily attracting, i.e., for $\theta_0 \in \mathsf{X}$,

$$\exists v_1, v_2, \dots \in \mathcal{O} = \mathrm{supp}\ p, \quad \lim F_k(\theta_0, v_1, \dots, v_k) = \theta^*.$$

If $\exists v_1, \dots, v_k \in \mathcal{O}$, such that

$$C_k = \mathrm{Jac}_{v_{1..k}} F_k(\theta^*, v_1, \dots, v_k)$$

is of rank n, then $\{\theta_k\}_k$ is an irreducible aperiodic T-chain.

Example

$$\theta_{k+1} = \theta_k + v_{k+1}$$

with $v_{k+1} \sim \mathcal{N}(0, I_d)$.

Example

$$\theta_{k+1} = \theta_k + v_{k+1}$$

with $v_{k+1} \sim \mathcal{N}(0, I_d)$.

Then: 0 is globally attracting.

Example

$$\theta_{k+1} = \theta_k + v_{k+1}$$

with $v_{k+1} \sim \mathcal{N}(0, I_d)$.

Then: 0 is globally attracting.

Proof.
Set $v_1 = -\theta_0 \in \mathrm{supp}\ p_{\mathcal{N}}$. □

Example

$$\theta_{k+1} = \theta_k + v_{k+1}$$

with $v_{k+1} \sim \mathcal{N}(0, I_d)$.

Then: 0 is globally attracting.

Proof.
Set $v_1 = -\theta_0 \in \mathrm{supp}\ p_{\mathcal{N}}$. □

Then: $\mathrm{rank}\,\mathrm{Jac}_v F(0, v) = n$

Example

$$\theta_{k+1} = \theta_k + v_{k+1}$$

with $v_{k+1} \sim \mathcal{N}(0, I_d)$.

Then: 0 is globally attracting.

Proof.
Set $v_1 = -\theta_0 \in \mathrm{supp}\ p_{\mathcal{N}}$. $\square$

Then: $\mathrm{rank}\,\mathrm{Jac}_v F(0, v) = n$

Proof.
$\mathrm{Jac}_v F(0, v) = I_n$ $\square$

Control model [Chotard & Auger 2019]

Consider
$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad \textbf{(CM(F))}$$

- $F \colon \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ **smooth** ($\mathcal{C}^\infty$)
- $\{v_{k+1}\}$ i.i.d. with a l.s.c. density p
- $\mathsf{X} \subset \mathbb{R}^n$, $\mathsf{V} \subset \mathbb{R}^p$ **open**.

Theorem
Suppose that θ^* is steadily attracting, i.e., for $\theta_0 \in \mathsf{X}$,

$$\exists v_1, v_2, \dots \in \mathcal{O}, \quad \lim F_k(\theta_0, v_1, \dots, v_k) = \theta^*.$$

If $\exists v_1, \dots, v_k \in \mathcal{O}$, such that

$$C_k = \mathrm{Jac}_{v_{1..k}} F_k(\theta^*, v_1, \dots, v_k)$$

is of rank n, then $\{\theta_k\}_k$ is an irreducible aperiodic T-chain.

Control model [Chotard & Auger 2019]

Consider
$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad \textbf{(CM(F))}$$

- $F : \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ is $\mathcal{C}^1$
- $\{v_{k+1}\}$ i.i.d. with a l.s.c. density p
- $\mathsf{X} \subset \mathbb{R}^n$, $\mathsf{V} \subset \mathbb{R}^p$ **open**.

Theorem
Suppose that θ^* is steadily attracting, i.e., for $\theta_0 \in \mathsf{X}$,

$$\exists v_1, v_2, \dots \in \mathcal{O}, \quad \lim F_k(\theta_0, v_1, \dots, v_k) = \theta^*.$$

If $\exists v_1, \dots, v_k \in \mathcal{O}$, such that

$$C_k = \mathrm{Jac}_{v_{1..k}} F_k(\theta^*, v_1, \dots, v_k)$$

is of rank n, then $\{\theta_k\}_k$ is an irreducible aperiodic T-chain.

Control model [Chotard & Auger 2019]

Consider

$$\theta_{k+1} = F(\theta_k, v_{k+1}) \qquad \textbf{(CM(F))}$$

- $F : \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ is $\mathcal{C}^1$
- $\{v_{k+1}\}$ such that $v_{k+1} \sim p_{\theta_k}$ l.s.c.
- $\mathsf{X} \subset \mathbb{R}^n$, $\mathsf{V} \subset \mathbb{R}^p$ **open**.

Theorem
Suppose that θ^* is steadily attracting, i.e., for $\theta_0 \in \mathsf{X}$,

$$\exists v_1, v_2, \dots \in \mathcal{O}, \quad \lim F_k(\theta_0, v_1, \dots, v_k) = \theta^*.$$

If $\exists v_1, \dots, v_k \in \mathcal{O}$, such that

$$C_k = \mathrm{Jac}_{v_{1..k}} F_k(\theta^*, v_1, \dots, v_k)$$

is of rank n, then $\{\theta_k\}_k$ is an irreducible aperiodic T-chain.

Control model [Chotard & Auger 2019]

Consider
$$\theta_{k+1} = F(\theta_k, \alpha(\theta_k, u_{k+1})) \qquad \textbf{(CM(F))}$$

- $F : \mathrm{X} \times \mathrm{V} \to \mathrm{X}$ is $\mathcal{C}^1$
- $\{v_{k+1}\}$ such that $v_{k+1} = \alpha(\theta_k, u_{k+1}) \sim p_{\theta_k}$ l.s.c.
- $\mathrm{X} \subset \mathbb{R}^n$, $\mathrm{V} \subset \mathbb{R}^p$ **open**.

Theorem
Suppose that θ^* is steadily attracting, i.e., for $\theta_0 \in \mathrm{X}$,

$$\exists v_1, v_2, \dots \in \mathcal{O}, \quad \lim F_k(\theta_0, v_1, \dots, v_k) = \theta^*.$$

If $\exists v_1, \dots, v_k \in \mathcal{O}$, such that

$$C_k = \mathrm{Jac}_{v_{1..k}} F_k(\theta^*, v_1, \dots, v_k)$$

is of rank n, then $\{\theta_k\}_k$ is an irreducible aperiodic T-chain.

Control model [Chotard & Auger 2019]

Consider $$\theta_{k+1} = F(\theta_k, \alpha(\theta_k, u_{k+1})) \qquad \textbf{(CM(F))}$$

- $F: \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ is $\mathcal{C}^1$
- $\{v_{k+1}\}$ such that $v_{k+1} = \alpha(\theta_k, u_{k+1}) \sim p_{\theta_k}$ l.s.c.
- $\mathsf{X} \subset \mathbb{R}^n$, $\mathsf{V} \subset \mathbb{R}^p$ **open**.

Theorem
Suppose that θ^* is steadily attracting, i.e., for $\theta_0 \in \mathsf{X}$,

$$\exists (v_1, v_2, \dots) \in \mathcal{O}_{\theta_0}^{\infty}, \quad \lim F_k(\theta_0, v_1, \dots, v_k) = \theta^*.$$

If $\exists (v_1, \dots, v_k) \in \mathcal{O}_{\theta^*}^{k}$, such that

$$C_k = \mathrm{Jac}_{v_{1..k}} F_k(\theta^*, v_1, \dots, v_k)$$

is of rank n, then $\{\theta_k\}_k$ is an irreducible aperiodic T-chain.

Application: Evolution Strategies (ES)

$$\text{Find } x^* \in \arg\min_{x \in \mathbb{R}^d} f(x) \tag{P}$$

Application: Evolution Strategies (ES)

$$\text{Find } x^* \in \arg\min_{x \in \mathbb{R}^d} f(x) \tag{P}$$

ES approximate x^* by $\mathcal{N}(m_k, \sigma_k^2 I_d)$

Application: Evolution Strategies (ES)

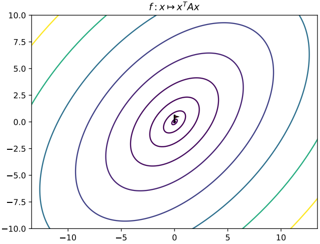
$$\text{Find } x^* \in \arg\min_{x \in \mathbb{R}^d} f(x) \tag{P}$$

ES approximate x^* by $\mathcal{N}(m_k, \sigma_k^2 I_d)$ by updating
$\theta_k = (m_k, \sigma_k) \in \mathbb{R}^d \times \mathbb{R}_{++}$.

Algorithm: ES with stepsize adaptation

Algorithm 1 ES

Goal: $\min_{x \in \mathbb{R}^d} f(x)$

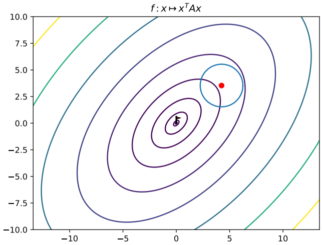


Algorithm: ES with stepsize adaptation

Algorithm 1 ES

Goal: $\min_{x \in \mathbb{R}^d} f(x)$

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$



Algorithm: ES with stepsize adaptation

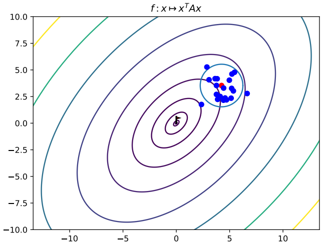
Algorithm 1 ES

Goal: $\min_{x \in \mathbb{R}^d} f(x)$

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$

For $k = 0, 1, 2, \ldots$:

1. $u_{k+1}^1, \ldots, u_{k+1}^\lambda \sim \mathcal{N}(0, I_d)$,
   $x_{k+1}^i = x_k + \sigma_k u_{k+1}^i \sim \mathcal{N}(m_k, \sigma_k^2 I_d)$



Algorithm: ES with stepsize adaptation

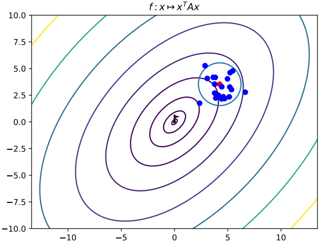
Algorithm 1 ES

Goal: $\min_{x \in \mathbb{R}^d} f(x)$

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$

For $k = 0, 1, 2, \ldots :$

1. $u^1_{k+1}, \ldots, u^\lambda_{k+1} \sim \mathcal{N}(0, I_d)$,
   $x^i_{k+1} = x_k + \sigma_k u^i_{k+1} \sim \mathcal{N}(m_k, \sigma_k^2 I_d)$
2. sort $f(x^i_{k+1})$:



Algorithm: ES with stepsize adaptation

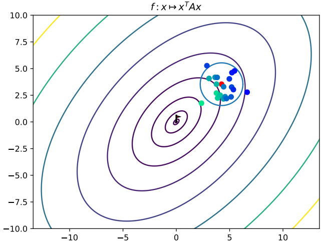
Algorithm 1 ES

Goal: $\min_{x \in \mathbb{R}^d} f(x)$

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$

For $k = 0, 1, 2, \ldots$:

1. $u^1_{k+1}, \ldots, u^\lambda_{k+1} \sim \mathcal{N}(0, I_d)$,
   $x^i_{k+1} = x_k + \sigma_k u^i_{k+1} \sim \mathcal{N}(m_k, \sigma_k^2 I_d)$
2. sort $f(x^i_{k+1})$:
   $f\left(x^{1:\lambda}_{k+1}\right) \leqslant \cdots \leqslant f\left(x^{\lambda:\lambda}_{k+1}\right)$



Algorithm: ES with stepsize adaptation

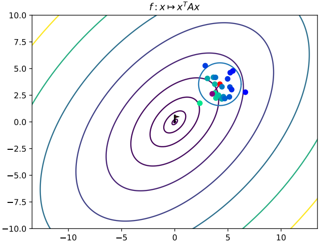
Algorithm 1 ES

Goal: $\min_{x \in \mathbb{R}^d} f(x)$

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$

For $k = 0, 1, 2, \ldots$:

1. $u^1_{k+1}, \ldots, u^\lambda_{k+1} \sim \mathcal{N}(0, I_d)$,
   $x^i_{k+1} = x_k + \sigma_k u^i_{k+1} \sim \mathcal{N}(m_k, \sigma_k^2 I_d)$
2. sort $f(x^i_{k+1})$:
   $f\left(x^{1:\lambda}_{k+1}\right) \leqslant \cdots \leqslant f\left(x^{\lambda:\lambda}_{k+1}\right)$
3. $m_{k+1} = \sum w_i x^{i:\lambda}_{k+1}$



Algorithm: ES with stepsize adaptation

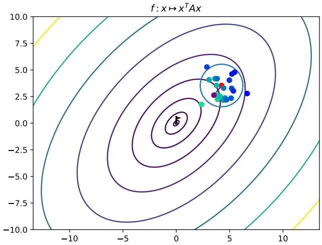
Algorithm 1 ES

Goal: $\min_{x \in \mathbb{R}^d} f(x)$

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$

For $k = 0, 1, 2, \ldots :$

1. $u^1_{k+1}, \ldots, u^\lambda_{k+1} \sim \mathcal{N}(0, I_d)$,
   $x^i_{k+1} = x_k + \sigma_k u^i_{k+1} \sim \mathcal{N}(m_k, \sigma_k^2 I_d)$
2. sort $f(x^i_{k+1})$:
   $f\left(x^{1:\lambda}_{k+1}\right) \leqslant \cdots \leqslant f\left(x^{\lambda:\lambda}_{k+1}\right)$
3. $m_{k+1} = \sum w_i x^{i:\lambda}_{k+1} = x_k + \sigma_k \sum w_i u^{i:\lambda}_{k+1}$



Algorithm: ES with stepsize adaptation

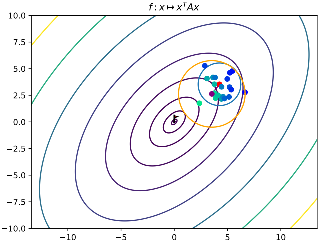
Algorithm 1 ES

Goal: $\min_{x \in \mathbb{R}^d} f(x)$

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$

For $k = 0, 1, 2, \ldots$:

1. $u_{k+1}^1, \ldots, u_{k+1}^\lambda \sim \mathcal{N}(0, I_d)$,
   $x_{k+1}^i = x_k + \sigma_k u_{k+1}^i \sim \mathcal{N}(m_k, \sigma_k^2 I_d)$
2. sort $f(x_{k+1}^i)$:
   $f\left(x_{k+1}^{1:\lambda}\right) \leqslant \cdots \leqslant f\left(x_{k+1}^{\lambda:\lambda}\right)$
3. $m_{k+1} = \sum w_i x_{k+1}^{i:\lambda} = x_k + \sigma_k \sum w_i u_{k+1}^{i:\lambda}$
4. $\log \sigma_{k+1} = \log \sigma_k + \dfrac{\left\|\sum w_i u_{k+1}^{i:\lambda}\right\|}{\mathbb{E}\left\|\sum w_i u_{k+1}^i\right\|} - 1$



Algorithm: ES with stepsize adaptation

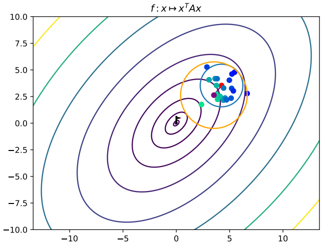
Algorithm 1 ES

Goal: $\min_{x \in \mathbb{R}^d} f(x)$

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$

For $k = 0, 1, 2, \ldots$:

1. $u_{k+1}^1, \ldots, u_{k+1}^\lambda \sim \mathcal{N}(0, I_d)$,
   $x_{k+1}^i = x_k + \sigma_k u_{k+1}^i \sim \mathcal{N}(m_k, \sigma_k^2 I_d)$
2. sort $f(x_{k+1}^i)$:
   $f\left(x_{k+1}^{1:\lambda}\right) \leqslant \cdots \leqslant f\left(x_{k+1}^{\lambda:\lambda}\right)$
3. $m_{k+1} = \sum w_i x_{k+1}^{i:\lambda} = x_k + \sigma_k \sum w_i u_{k+1}^{i:\lambda}$
4. $\log \sigma_{k+1} = \log \sigma_k + \frac{\left\|\sum w_i u_{k+1}^{i:\lambda}\right\|^2}{\mathbb{E}\left\|\sum w_i u_{k+1}^i\right\|^2} - 1$



Algorithm: ES with stepsize adaptation

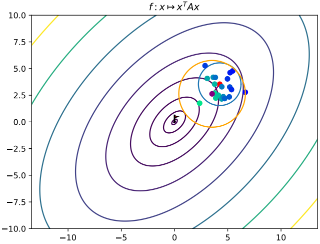
Algorithm 1 ES

Goal: $\min_{x \in \mathbb{R}^d} f(x)$

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$

For $k = 0, 1, 2, \ldots$:

1. $u^1_{k+1}, \ldots, u^\lambda_{k+1} \sim \mathcal{N}(0, I_d)$,
   $x^i_{k+1} = x_k + \sigma_k u^i_{k+1} \sim \mathcal{N}(m_k, \sigma_k^2 I_d)$
2. sort $f(x^i_{k+1})$:
   $f\left(x^{1:\lambda}_{k+1}\right) \leqslant \cdots \leqslant f\left(x^{\lambda:\lambda}_{k+1}\right)$
3. $m_{k+1} = \sum w_i x^{i:\lambda}_{k+1} = x_k + \sigma_k \sum w_i u^{i:\lambda}_{k+1}$
4. $\sigma_{k+1} = \sigma_k \times \exp\left(\frac{\left\| \sum w_i u^{i:\lambda}_{k+1} \right\|^2}{\mathbb{E} \left\| \sum w_i u^i_{k+1} \right\|^2} - 1 \right)$



Convergence via analysis of CM(F)

$$\theta_k = (m_k, \sigma_k) \underset{k\to\infty}{\longrightarrow} (x^*, 0)$$

Convergence via analysis of CM(F)

$$\theta_k = (m_k, \sigma_k) \underset{k\to\infty}{\longrightarrow} (x^*, 0)$$

$\Rightarrow$ no stationary measure

Convergence via analysis of CM(F)

$$\theta_k = (m_k, \sigma_k) \underset{k\to\infty}{\longrightarrow} (x^*, 0)$$

$\Rightarrow$ no stationary measure

$$z_k \stackrel{\text{def}}{=} \frac{m_k - x^*}{\sigma_k}$$

Convergence via analysis of CM(F)

$$\theta_k = (m_k, \sigma_k) \underset{k\to\infty}{\longrightarrow} (x^*, 0)$$

$\Rightarrow$ no stationary measure

$$z_k \stackrel{\text{def}}{=} \frac{m_k - x^*}{\sigma_k}$$

$$z_{k+1} = F(z_k, v_{k+1})$$

Convergence via analysis of CM(F)

$$\theta_k = (m_k, \sigma_k) \underset{k\to\infty}{\longrightarrow} (x^*, 0)$$

$\Rightarrow$ no stationary measure

$$z_k \overset{\text{def}}{=} \frac{m_k - x^*}{\sigma_k}$$

$$\begin{aligned} z_{k+1} &= F(z_k, v_{k+1}) \\ &= \exp\left(1 - \frac{\|w^T v_{k+1}\|^2}{\mathbb{E}\|w^T u_{k+1}\|^2}\right) \times (z_k + w^T v_{k+1}) \end{aligned}$$

Convergence via analysis of CM(F)

$$\theta_k = (m_k, \sigma_k) \underset{k\to\infty}{\longrightarrow} (x^*, 0)$$

$\Rightarrow$ no stationary measure

$$z_k \stackrel{\text{def}}{=} \frac{m_k - x^*}{\sigma_k}$$

$$\begin{aligned} z_{k+1} &= F(z_k, v_{k+1}) \\ &= \exp\left(1 - \frac{\|w^T v_{k+1}\|^2}{\mathbb{E}\|w^T u_{k+1}\|^2}\right) \times (z_k + w^T v_{k+1}) \end{aligned}$$

with $v_{k+1} = (u_{k+1}^{1:\lambda}, \ldots, u_{k+1}^{\lambda:\lambda})$.

Control model [Chotard & Auger 2019]

Consider
$$\theta_{k+1} = F(\theta_k, \alpha(\theta_k, u_{k+1})) \qquad \textbf{(CM(F))}$$

- $F : \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ is $\mathcal{C}^1$
- $\{v_{k+1}\}$ such that $v_{k+1} = \alpha(\theta_k, u_{k+1}) \sim p_{\theta_k}$ l.s.c.
- $\mathsf{X} \subset \mathbb{R}^n$, $\mathsf{V} \subset \mathbb{R}^p$ **open**.

Theorem
Suppose that θ^* is steadily attracting, i.e., for $\theta_0 \in \mathsf{X}$,

$$\exists (v_1, v_2, \dots) \in \mathcal{O}_{\theta_0}^{\infty}, \quad \lim F_k(\theta_0, v_1, \dots, v_k) = \theta^*.$$

If $\exists (v_1, \dots, v_k) \in \mathcal{O}_{\theta^*}^{k}$, such that

$$C_k = \mathrm{Jac}_{v_{1..k}} F_k(\theta^*, v_1, \dots, v_k)$$

is of rank n, then $\{\theta_k\}_k$ is an irreducible aperiodic T-chain.

Irreducibility of ES

$$
\begin{aligned}
z_{k+1} &= F(z_k, v_{k+1}) \\
&= \exp\left(1 - \frac{\|w^T v_{k+1}\|^2}{\mathbb{E}\|w^T u_{k+1}\|^2}\right) \times (z_k + w^T v_{k+1})
\end{aligned}
$$

Irreducibility of ES

$$\begin{aligned} z_{k+1} &= F(z_k, v_{k+1}) \\ &= \exp\left(1 - \frac{\|w^T v_{k+1}\|^2}{\mathbb{E}\|w^T u_{k+1}\|^2}\right) \times (z_k + w^T v_{k+1}) \end{aligned}$$

Then: 0 is globally attracting.

Irreducibility of ES

$$\begin{aligned} z_{k+1} &= F(z_k, v_{k+1}) \\ &= \exp\left(1 - \frac{\|w^T v_{k+1}\|^2}{\mathbb{E}\|w^T u_{k+1}\|^2}\right) \times (z_k + w^T v_{k+1}) \end{aligned}$$

Then: 0 is globally attracting.

Proof.

□

Irreducibility of ES

$$
\begin{aligned}
z_{k+1} &= F(z_k, v_{k+1}) \\
&= \exp\left(1 - \frac{\|w^T v_{k+1}\|^2}{\mathbb{E}\|w^T u_{k+1}\|^2}\right) \times (z_k + w^T v_{k+1})
\end{aligned}
$$

Then: 0 is globally attracting.

Proof.

If $v_1 = (-z_0, \ldots, -z_0)$,

□

Irreducibility of ES

$$
\begin{aligned}
z_{k+1} &= F(z_k, v_{k+1}) \\
&= \exp\left(1 - \frac{\|w^T v_{k+1}\|^2}{\mathbb{E}\|w^T u_{k+1}\|^2}\right) \times (z_k + w^T v_{k+1})
\end{aligned}
$$

Then: 0 is globally attracting.

Proof.

If $v_1 = (-z_0, \ldots, -z_0)$,
then, $z_1 = \exp(\ldots) \times (z_0 - \sum w_i z_0) = 0$. $\square$

Irreducibility of ES

$$\begin{aligned} z_{k+1} &= F(z_k, v_{k+1}) \\ &= \exp\left(1 - \frac{\|w^T v_{k+1}\|^2}{\mathbb{E}\|w^T u_{k+1}\|^2}\right) \times (z_k + w^T v_{k+1}) \end{aligned}$$

Then: 0 is globally attracting.

Proof.
If $v_1 = (-z_0, \dots, -z_0)$,
then, $z_1 = \exp(\dots) \times (z_0 - \sum w_i z_0) = 0$. □

Then: $\operatorname{rank} \operatorname{Jac}_v F(0,0) = n$.

Proof.
□

Irreducibility of ES

$$
\begin{aligned}
z_{k+1} &= F(z_k, v_{k+1}) \\
&= \exp\left(1 - \frac{\|w^T v_{k+1}\|^2}{\mathbb{E}\|w^T u_{k+1}\|^2}\right) \times (z_k + w^T v_{k+1})
\end{aligned}
$$

Then: 0 is globally attracting.

Proof.

If $v_1 = (-z_0, \ldots, -z_0)$,
then, $z_1 = \exp(\ldots) \times (z_0 - \sum w_i z_0) = 0$. □

Then: $\operatorname{rank} \operatorname{Jac}_v F(0,0) = n$.

Proof.
$F(0, 0+h) = \exp(1 + O(h)) \times w^T h$ □

Irreducibility of ES

$$
\begin{aligned}
z_{k+1} &= F(z_k, v_{k+1}) \\
&= \exp\left(1 - \frac{\|w^T v_{k+1}\|^2}{\mathbb{E}\|w^T u_{k+1}\|^2}\right) \times (z_k + w^T v_{k+1})
\end{aligned}
$$

Then: 0 is globally attracting.

Proof.

If $v_1 = (-z_0, \dots, -z_0)$,
then, $z_1 = \exp(\dots) \times (z_0 - \sum w_i z_0) = 0$. □

Then: $\operatorname{rank} \operatorname{Jac}_v F(0,0) = n$.

Proof.
$F(0, 0+h) = \exp(1 + O(h)) \times w^T h = w^T h + o(h)$. □

Convergence

1. $\{z_k\}_{k\in\mathbb{N}}$ is **irreducible** and **aperiodic**

Convergence

1. $\{z_k\}_{k\in\mathbb{N}}$ is **irreducible** and **aperiodic**
2. we have a **drift** condition with

$$V(z) = \|z\|^\alpha$$

Convergence

1. $\{z_k\}_{k\in\mathbb{N}}$ is **irreducible** and **aperiodic**

2. we have a **drift** condition with

$$V(z) = \|z\|^\alpha$$

$\Rightarrow \{z_k\}_{k\in\mathbb{N}}$ admits a **unique invariant probability measure** π

Convergence

1. $\{z_k\}_{k\in\mathbb{N}}$ is **irreducible** and **aperiodic**
2. we have a **drift** condition with

$$V(z) = \|z\|^\alpha$$

$\Rightarrow \{z_k\}_{k\in\mathbb{N}}$ admits a **unique invariant probability measure** π and

$$\lim_{t\to\infty} \frac{1}{t} \log \frac{\|m_t - x^*\|}{\|m_0 - x^*\|}$$

$$= \lim_{t\to\infty} \frac{1}{t} \sum_{k=0}^{t-1} \log \frac{\|z_{k+1}\|}{\|z_k\|} - \log \exp \left(\frac{\left\|\sum w_i u_{k+1}^{i:\lambda}\right\|^2}{\mathbb{E}\left\|\sum w_i u_{k+1}^{i}\right\|^2} - 1 \right)$$

Convergence

1. $\{z_k\}_{k\in\mathbb{N}}$ is **irreducible** and **aperiodic**
2. we have a **drift** condition with

$$V(z) = \|z\|^\alpha$$

$\Rightarrow$ $\{z_k\}_{k\in\mathbb{N}}$ admits a **unique invariant probability measure** π and

$$\begin{aligned}
&\lim_{t\to\infty} \frac{1}{t} \log \frac{\|m_t - x^*\|}{\|m_0 - x^*\|} \\
&\quad = \lim_{t\to\infty} \frac{1}{t} \sum_{k=0}^{t-1} \log \frac{\|z_{k+1}\|}{\|z_k\|} - \log \exp\left(\frac{\left\|\sum w_i u_{k+1}^{i:\lambda}\right\|^2}{\mathbb{E}\left\|\sum w_i u_{k+1}^{i}\right\|^2} - 1 \right) \\
&\quad = -\mathbb{E}_\pi \left[\frac{\left\|\sum w_i u^{i:\lambda}\right\|^2}{\mathbb{E}\left\|\sum w_i u^{i}\right\|^2} - 1 \;\middle|\; u^1, \ldots, u^\lambda \sim \mathcal{N}_d \right]
\end{aligned}$$

Convergence

1. $\{z_k\}_{k\in\mathbb{N}}$ is **irreducible** and **aperiodic**
2. we have a **drift** condition with

$$V(z) = \|z\|^\alpha$$

$\Rightarrow$ $\{z_k\}_{k\in\mathbb{N}}$ admits a **unique invariant probability measure** π and

$$\lim_{t\to\infty} \frac{1}{t} \log \frac{\|m_t - x^*\|}{\|m_0 - x^*\|}$$

$$= \lim_{t\to\infty} \frac{1}{t} \sum_{k=0}^{t-1} \log \frac{\|z_{k+1}\|}{\|z_k\|} - \log \exp\left(\frac{\left\|\sum w_i u^{i:\lambda}_{k+1}\right\|^2}{\mathbb{E}\left\|\sum w_i u^{i}_{k+1}\right\|^2} - 1 \right)$$

$$= -\mathbb{E}_\pi \left[\frac{\left\|\sum w_i u^{i:\lambda}\right\|^2}{\mathbb{E}\left\|\sum w_i u^{i}\right\|^2} - 1 \;\middle|\; u^1, \ldots, u^\lambda \sim \mathcal{N}_d \right]$$

Conclusion: ES converges *linearly* to x^*

Control model, [G., Auger & Durmus, soon submitted]

$$\theta_{k+1} = F(\theta_k, \alpha(\theta_k, u_{k+1})) \qquad \textbf{(CM(F))}$$

- $F\colon \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ is $\mathcal{C}^1$
- $\{v_{k+1}\}$ such that $v_{k+1} = \alpha(\theta_k, u_{k+1}) \sim p_{\theta_k}$ l.s.c.
- $\mathsf{X} \subset \mathbb{R}^n$, $\mathsf{V} \subset \mathbb{R}^p$ **open**.

Theorem
Suppose that θ^* is steadily attracting, i.e., for $\theta_0 \in \mathsf{X}$,

$$\exists (v_1, v_2, \dots) \in \mathcal{O}_{\theta_0}^{\infty}, \quad \lim F_k(\theta_0, v_1, \dots, v_k) = \theta^*.$$

If $\exists (v_1, \dots, v_k) \in \mathcal{O}_{\theta^*}^{k}$, such that

$$C_k = \mathrm{Jac}_{v_{1..k}} F_k(\theta^*, v_1, \dots, v_k)$$

is of rank n, then $\{\theta_k\}_k$ is an irreducible aperiodic T-chain.

Control model, [G., Auger & Durmus, soon submitted]

$$\theta_{k+1} = F(\theta_k, \alpha(\theta_k, u_{k+1})) \qquad \textbf{(CM(F))}$$

- $F : \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ is **locally Lipschitz**
- $\{v_{k+1}\}$ such that $v_{k+1} = \alpha(\theta_k, u_{k+1}) \sim p_{\theta_k}$ l.s.c.
- $\mathsf{X} \subset \mathbb{R}^n$, $\mathsf{V} \subset \mathbb{R}^p$ **open**.

Theorem
Suppose that θ^* is steadily attracting, i.e., for $\theta_0 \in \mathsf{X}$,

$$\exists (v_1, v_2, \dots) \in \mathcal{O}_{\theta_0}^{\infty}, \quad \lim F_k(\theta_0, v_1, \dots, v_k) = \theta^*.$$

If $\exists (v_1, \dots, v_k) \in \mathcal{O}_{\theta^*}^{k}$, such that

$$C_k = \mathrm{Jac}_{v_{1..k}} F_k(\theta^*, v_1, \dots, v_k)$$

is of rank n, then $\{\theta_k\}_k$ is an irreducible aperiodic T-chain.

Control model, [G., Auger & Durmus, soon submitted]

$$\theta_{k+1} = F(\theta_k, \alpha(\theta_k, u_{k+1})) \qquad \textbf{(CM(F))}$$

- $F : \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ is **locally Lipschitz**
- $\{v_{k+1}\}$ such that $v_{k+1} = \alpha(\theta_k, u_{k+1}) \sim p_{\theta_k}$ l.s.c.
- $\mathsf{X} \subset \mathbb{R}^n$, $\mathsf{V} \subset \mathbb{R}^p$ **open**.

Theorem
Suppose that θ^* is steadily attracting, i.e., for $\theta_0 \in \mathsf{X}$,

$$\exists (v_1, v_2, \dots) \in \mathcal{O}_{\theta_0}^{\infty}, \quad \lim F_k(\theta_0, v_1, \dots, v_k) = \theta^*.$$

If $\exists (v_1, \dots, v_k) \in \mathcal{O}_{\theta^*}^{k}$, such that any

$$C_k \in \mathrm{Clarke}_{v_{1..k}} F_k(\theta^*, v_1, \dots, v_k)$$

is of rank n, then $\{\theta_k\}_k$ is an irreducible aperiodic T-chain.

Control model, [G., Auger & Durmus, soon submitted]

$$\theta_{k+1} = F(\theta_k, \alpha(\theta_k, u_{k+1})) \qquad \textbf{(CM(F))}$$

- $F \colon \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ is **locally Lipschitz**
- $\{v_{k+1}\}$ such that $v_{k+1} = \alpha(\theta_k, u_{k+1}) \sim p_{\theta_k}$ l.s.c.
- X and V are **manifolds**

Theorem
Suppose that θ^* is steadily attracting, i.e., for $\theta_0 \in \mathsf{X}$,

$$\exists (v_1, v_2, \dots) \in \mathcal{O}_{\theta_0}^{\infty}, \quad \lim F_k(\theta_0, v_1, \dots, v_k) = \theta^*.$$

If $\exists (v_1, \dots, v_k) \in \mathcal{O}_{\theta^*}^{k}$, such that any

$$C_k \in \mathrm{Clarke}_{v_{1..k}} F_k(\theta^*, v_1, \dots, v_k)$$

is of rank n, then $\{\theta_k\}_k$ is an irreducible aperiodic T-chain.

Algorithm: ES with covariance matrix adapation

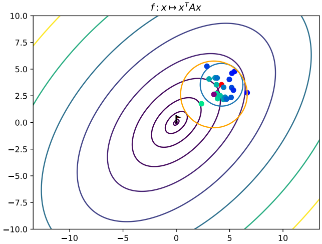
Algorithm 1 ES

Goal: $\min_{x \in \mathbb{R}^d} f(x)$

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$

For $k = 0, 1, 2, \dots$:

1. $u^1_{k+1}, \dots, u^\lambda_{k+1} \sim \mathcal{N}(0, I_d)$,
   $x^i_{k+1} = x_k + \sigma_k u^i_{k+1} \sim \mathcal{N}(m_k, \sigma_k^2 I_d)$
2. sort $f(x^i_{k+1})$:
   $f\left(x^{1:\lambda}_{k+1}\right) \leqslant \cdots \leqslant f\left(x^{\lambda:\lambda}_{k+1}\right)$
3. $m_{k+1} = \sum w_i x^{i:\lambda}_{k+1} = x_k + \sigma_k \sum w_i u^{i:\lambda}_{k+1}$
4. $\sigma_{k+1} = \sigma_k \times \exp\left(\frac{\left\| \sum w_i u^{i:\lambda}_{k+1} \right\|^2}{\mathbb{E}\left\| \sum w_i u^i_{k+1} \right\|^2} - 1 \right)$



Algorithm: ES with covariance matrix adapation

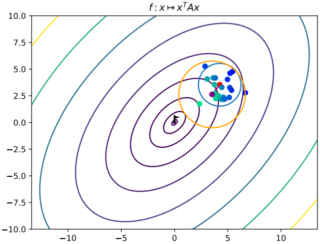
Algorithm 1 ES

Goal: $\min_{x \in \mathbb{R}^d} f(x)$

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$

For $k = 0, 1, 2, \ldots$:

1. $u^1_{k+1}, \ldots, u^\lambda_{k+1} \sim \mathcal{N}(0, I_d)$,
   $x^i_{k+1} = x_k + \sigma_k u^i_{k+1} \sim \mathcal{N}(m_k, \sigma_k^2 I_d)$
2. sort $f(x^i_{k+1})$:
   $f\left(x^{1:\lambda}_{k+1}\right) \leqslant \cdots \leqslant f\left(x^{\lambda:\lambda}_{k+1}\right)$
3. $m_{k+1} = \sum w_i x^{i:\lambda}_{k+1} = x_k + \sigma_k \sum w_i u^{i:\lambda}_{k+1}$
4. $\sigma_{k+1} = \sigma_k \times \exp\left(\frac{\left\|\sum w_i u^{i:\lambda}_{k+1}\right\|}{\mathbb{E}\left\|\sum w_i u^i_{k+1}\right\|} - 1\right)$



Algorithm: ES with covariance matrix adapation

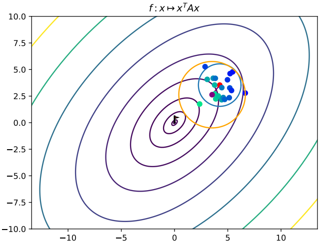
Algorithm 1 CMA-ES

Goal: $\min_{x \in \mathbb{R}^d} f(x)$

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$, $C_0 \in \mathcal{S}^d_{++}$

For $k = 0, 1, 2, \ldots$:

1. $u^1_{k+1}, \ldots, u^\lambda_{k+1} \sim \mathcal{N}(0, I_d)$,
   $x^i_{k+1} = x_k + \sigma_k \sqrt{C_k} u^i_{k+1}$
2. sort $f(x^i_{k+1})$:
   $f\left(x^{1:\lambda}_{k+1}\right) \leqslant \cdots \leqslant f\left(x^{\lambda:\lambda}_{k+1}\right)$
3. $m_{k+1} = \sum w_i x^{i:\lambda}_{k+1} = x_k + \sigma_k \sqrt{C_k} \sum w_i u^{i:\lambda}_{k+1}$
4. $\sigma_{k+1} = \sigma_k \times \exp\left(\frac{\left\|\sum w_i u^{i:\lambda}_{k+1}\right\|}{\mathbb{E}\left\|\sum w_i u^i_{k+1}\right\|} - 1 \right)$



Algorithm: ES with covariance matrix adapation

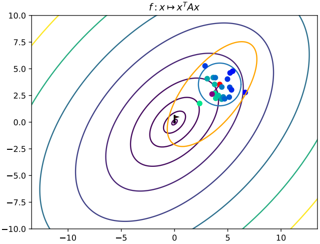
Algorithm 1 CMA-ES

Goal: $\min_{x \in \mathbb{R}^d} f(x)$

Given: $m_0 \in \mathbb{R}^d$, $\sigma_0 > 0$, $C_0 \in \mathcal{S}^d_{++}$

For $k = 0, 1, 2, \ldots$:

1. $u^1_{k+1}, \ldots, u^\lambda_{k+1} \sim \mathcal{N}(0, I_d)$,
   $x^i_{k+1} = x_k + \sigma_k \sqrt{C_k} u^i_{k+1}$
2. sort $f(x^i_{k+1})$:
   $f\left(x^{1:\lambda}_{k+1}\right) \leqslant \cdots \leqslant f\left(x^{\lambda:\lambda}_{k+1}\right)$
3. $m_{k+1} = \sum w_i x^{i:\lambda}_{k+1} = x_k + \sigma_k \sqrt{C_k} \sum w_i u^{i:\lambda}_{k+1}$
4. $\sigma_{k+1} = \sigma_k \times \exp\left(\frac{\left\|\sum w_i u^{i:\lambda}_{k+1}\right\|}{\mathbb{E}\left\|\sum w_i u^i_{k+1}\right\|} - 1\right)$
5. $C_{k+1} = (1-c) C_k + c \sum w_i \left[\sqrt{C_k} u^{i:\lambda}_{t+1}\right] \left[\sqrt{C_k} u^{i:\lambda}_{t+1}\right]^T$



CMA-ES follows CM(F)

$$\theta_k = (m_k, \sigma_k, C_k) \underset{k \to \infty}{\longrightarrow} (x^*, 0, 0)$$

CMA-ES follows CM(F)

$$\theta_k = (m_k, \sigma_k, C_k) \underset{k \to \infty}{\longrightarrow} (x^*, 0, 0)$$

$\Rightarrow$ no stationary measure

CMA-ES follows CM(F)

$$\theta_k = (m_k, \sigma_k, C_k) \underset{k\to\infty}{\longrightarrow} (x^*, 0, 0)$$

$\Rightarrow$ no stationary measure

$$z_k \stackrel{\text{def}}{=} \frac{m_k - x^*}{\sigma_k}$$

CMA-ES follows CM(F)

$$\theta_k = (m_k, \sigma_k, C_k) \underset{k\to\infty}{\longrightarrow} (x^*, 0, 0)$$

$\Rightarrow$ no stationary measure

$$z_k \overset{\text{def}}{=} \frac{m_k - x^*}{\sigma_k \det^{1/2d}(C_k)}$$

CMA-ES follows CM(F)

$$\theta_k = (m_k, \sigma_k, C_k) \underset{k\to\infty}{\longrightarrow} (x^*, 0, 0)$$

$\Rightarrow$ no stationary measure

$$z_k \stackrel{\text{def}}{=} \frac{m_k - x^*}{\sigma_k \det^{1/2d}(C_k)}$$

$$\Sigma_k \stackrel{\text{def}}{=} \frac{C_k}{\det^{1/d}(C_k)}$$

CMA-ES follows CM(F)

$$\theta_k = (m_k, \sigma_k, C_k) \underset{k\to\infty}{\longrightarrow} (x^*, 0, 0)$$

$\Rightarrow$ no stationary measure

$$z_k \overset{\text{def}}{=} \frac{m_k - x^*}{\sigma_k \det^{1/2d}(C_k)}$$

$$\Sigma_k \overset{\text{def}}{=} \frac{C_k}{\det^{1/d}(C_k)} \in \det^{-1}(\{1\})$$

CMA-ES follows CM(F)

$$\theta_k = (m_k, \sigma_k, C_k) \underset{k\to\infty}{\longrightarrow} (x^*, 0, 0)$$

$\Rightarrow$ no stationary measure

$$z_k \overset{\text{def}}{=} \frac{m_k - x^*}{\sigma_k \det^{1/2d}(C_k)}$$

$$\Sigma_k \overset{\text{def}}{=} \frac{C_k}{\det^{1/d}(C_k)} \in \det^{-1}(\{1\})$$

$$(z_{k+1}, \Sigma_{k+1}) = F((z_k, \Sigma_k), v_{k+1})$$

CMA-ES follows CM(F)

$$\theta_k = (m_k, \sigma_k, C_k) \underset{k\to\infty}{\longrightarrow} (x^*, 0, 0)$$

$\Rightarrow$ no stationary measure

$$z_k \overset{\text{def}}{=} \frac{m_k - x^*}{\sigma_k \det^{1/2d}(C_k)}$$

$$\Sigma_k \overset{\text{def}}{=} \frac{C_k}{\det^{1/d}(C_k)} \in \det{}^{-1}(\{1\})$$

$$(z_{k+1}, \Sigma_{k+1}) = F((z_k, \Sigma_k), v_{k+1})$$

with $v_{k+1} = (u_{k+1}^{1:\lambda}, \ldots, u_{k+1}^{\lambda:\lambda})$.

Control model, [G., Auger & Durmus, soon submitted]

$$\theta_{k+1} = F(\theta_k, \alpha(\theta_k, u_{k+1})) \qquad \textbf{(CM(F))}$$

- $F : \mathsf{X} \times \mathsf{V} \to \mathsf{X}$ is **locally Lipschitz**
- $\{v_{k+1}\}$ such that $v_{k+1} = \alpha(\theta_k, u_{k+1}) \sim p_{\theta_k}$ l.s.c.
- X and V are **manifolds**

Theorem
Suppose that θ^* is steadily attracting, i.e., for $\theta_0 \in \mathsf{X}$,

$$\exists (v_1, v_2, \dots) \in \mathcal{O}^{\infty}_{\theta_0}, \quad \lim F_k(\theta_0, v_1, \dots, v_k) = \theta^*.$$

If $\exists (v_1, \dots, v_k) \in \mathcal{O}^{k}_{\theta^*}$, such that any

$$C_k \in \mathrm{Clarke}_{v_{1..k}} F_k(\theta^*, v_1, \dots, v_k)$$

is of rank n, then $\{\theta_k\}_k$ is an irreducible aperiodic T-chain.

Consequence

Corollary
Under additional assumptions on the objective function f,

$$(z_k, \Sigma_k)_{k \in \mathbb{N}}$$

is a irreducible, aperiodic T-chain.

Drift

Then (z_k, Σ_k) is positive recurrent and follows a LLN if there exists a drift $V\colon \mathsf{X} \to [0, +\infty]$ such that

$$\mathbb{E}\left[V(z_1, \Sigma_1)\right] \leqslant (1 - \varepsilon) \times V(z_0, \Sigma_0) + b \times \mathbf{1}_{(z_0, \Sigma_0) \in K}$$

Theorem (Drift for the normalized chain)
When minimizing a **spherical** *function* $f : x \mapsto g\left(x^T x\right)$ *(*$g : \mathbb{R} \to \mathbb{R}$ *increasing), then the irreducible, aperiodic Markov chain* $(z_k, \Sigma_k)_{k \in \mathbb{N}}$ *satisfies a Foster-Lyapunov condition with the potential defined by*

$$V(z, \Sigma) = \alpha \times \frac{\|\Sigma z\|^2}{\lambda_{\max}(\Sigma)^2} + \beta \times \lambda_{\max}(\Sigma)$$

Theorem (Drift for the normalized chain)
When minimizing a **spherical** *function* $f : x \mapsto g\left(x^T x\right)$ *($g : \mathbb{R} \to \mathbb{R}$ increasing), then the irreducible, aperiodic Markov chain* $(z_k, \Sigma_k)_{k \in \mathbb{N}}$ *satisfies a Foster-Lyapunov condition with the potential defined by*

$$V(z, \Sigma) = \alpha \times \frac{\|\Sigma z\|^2}{\lambda_{\max}(\Sigma)^2} + \beta \times \lambda_{\max}(\Sigma)$$

This can be generalized to when minimizing **ellipsoid** functions $f(x) = g(x^T H x)$ using the **affine-invariance** of CMA-ES.

Affine-Invariance

$$
\begin{array}{ccc}
(m_0, C_0) & \xrightarrow{\min f(x)} & (m_1, C_1) \\
\Big\downarrow \Psi & & \Big\uparrow \Psi^{-1} \\
(m'_0, C'_0) & \xrightarrow{\min f(Bx + b)} & (m'_1, C'_1)
\end{array}
$$

Conclusion

- $(z_k, \Sigma_k)_{k \in \mathbb{N}}$ is **irreducible** and **aperiodic** ;
- it is **positive recurrent** ;
- we deduce a LLN and the **convergence** of CMA-ES.

Thank you!

Scaling-invariant functions