Evolution Strategies for Constrained Optimization

Working meeting “Dealing with stochastics in optimization problems”, IHP, Paris

Asma Atamna

(Joint work with Anne Auger & Nikolaus Hansen)

Inria, LRI, Université Paris-Saclay

May 13, 2016

Outline

Context

- Constrained optimization problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \text{ s.t. } g(\mathbf{x}) \leq 0$$

$f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \to \mathbb{R}, n \in \mathbb{N}$

- Black-box setting

Contributions

- General framework for building stochastic derivative-free optimization (DFO) algorithm for constrained optimization
  - From stochastic DFO algorithm for unconstrained optimization
  - Using augmented Lagrangian constraint handling approach
- Algorithm: $(\mu/\mu_w, \lambda)$-MSR-CMA-ES with adaptive augmented Lagrangian
- Theoretical framework for analyzing linear convergence

Outline

Stochastic derivative-free optimization algorithm

$$\min_{\mathbf{x}} f(\mathbf{x}), f : \mathbb{R}^n \to \mathbb{R}, n \in \mathbb{N}$$

- Sequence $(\mathbf{s}_t)_{t\in\mathbb{N}}$ of states $\mathbf{s}_t$
- Transition function $\mathcal{F}^f$

$$\mathbf{s}_{t+1} = \mathcal{F}^f(\mathbf{s}_t, \mathbf{U}_{t+1})$$

  $\mathbf{U}_{t+1} = [\mathbf{U}^1_{t+1}, \cdots, \mathbf{U}^\lambda_{t+1}]$, $\mathbf{U}^k_{t+1}$ i.i.d. random vectors
- $\mathbf{s}_t$ includes a vector $\mathbf{X}_t \in \mathbb{R}^n$ representing the favorite solution at iteration t
- From $\mathbf{s}_t$ and $\mathbf{U}_{t+1}$, sample λ candidate solutions $\mathbf{X}^k_{t+1}$
- $\mathbf{X}^k_{t+1}$ evaluated on f

Evolution strategies (ES)

- Stochastic DFO algorithms
- λ normally distributed candidate solutions (offspring)

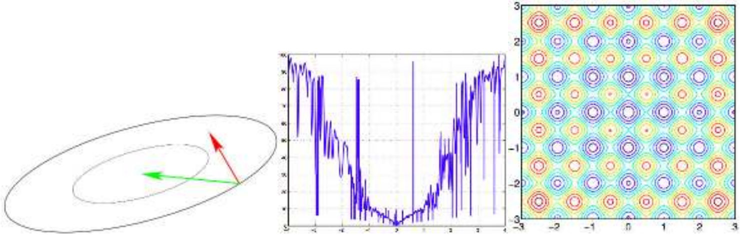
$$\mathbf{X}_{t+1}^{k} = \mathbf{X}_t + \sigma_t \mathbf{U}_{t+1}^{k}, \mathbf{U}_{t+1}^{k} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_t)$$

  - $\mathbf{X}_t$: current solution (mean vector)
  - $\sigma_t \in \mathbb{R}_{>}^{+}$: step-size
  - $\mathbf{C}_t$: covariance matrix
- $(\mu/\mu_w, \lambda)$-ES: Recombine μ best offspring (parents)

$$\mathbf{X}_{t+1} = \sum_{k=1}^{\mu} w_k \mathbf{X}_{t+1}^{k:\lambda}$$

Covariance matrix adaptation ES (CMA-ES) [Hansen et al.'01]

- State-of-the-art ES
- Adapt $\mathbf{C}_t$ to increase likelihood of successful solutions
- Efficiently tackles ill-conditioned, noisy, non-separable functions
- Shows linear convergence on unconstrained optimization problems



(a) Ill-conditioned (b) Noisy (c) Non-separable

Outline

Linear Convergence

- Central property in unconstrained optimization
- Preserve linear convergence in constrained optimization
- Minimal requirement for evolutionary algorithm for constrained optimization [Arnold et al.'15]
  - Converge linearly on convex quadratic functions with one linear constraint $g(\mathbf{x}) \leq 0$, $g(\mathbf{x}) = \mathbf{x}^{\mathsf{T}}\mathbf{b} + c$, $\mathbf{b} \in \mathbb{R}^n$

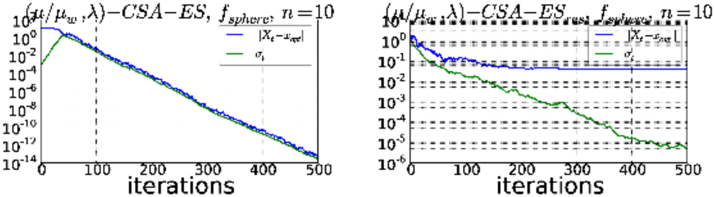


Figure: Left: single run of $(\mu/\mu_w, \lambda)$-CSA-ES on unconstrained f_{sphere}. Right: single run of $(\mu/\mu_w, \lambda)$-CSA-ES with resampling on linearly constrained f_{sphere}

Outline

- Constraint handling methods that transform a constrained optimization problem into an unconstrained one
- Augmented Lagrangian: a combination of a Lagrangian $\mathcal{L} : \mathbb{R}^{n+1} \to \mathbb{R}$ and a penalty function

$\to$ avoid the shortcomings of penalty function methods

Lagrangian associated to our constrained problem

$$\mathcal{L}(\mathbf{x}, \gamma) = f(\mathbf{x}) + \gamma g(\mathbf{x})$$

Karush-Kuhn-Tucker (KKT) stationarity condition

Consider the problem of minimizing the objective function $f : \mathbb{R}^n \to \mathbb{R}$ s.t. the constraint $g(\mathbf{x}) \leq 0$, where $g : \mathbb{R}^n \to \mathbb{R}$. If $\mathbf{x}^*$ is a local optimum that satisfies some regularity conditions, then there exists a non-negative constant γ^*, called the Lagrange multiplier, such that

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \gamma^* \nabla_{\mathbf{x}} g(\mathbf{x}^*) = \mathbf{0}$$

Augmented Lagrangian Methods

Considered Augmented Lagrangian

$$\underbrace{h(\mathbf{x},\gamma,\omega)}_{\text{augmented Lagrangian}} = f(\mathbf{x}) + \begin{cases} \gamma g(\mathbf{x}) + \frac{\omega}{2} g^2(\mathbf{x}) & \text{if } \gamma + \omega g(\mathbf{x}) \geq 0 \\ -\frac{\gamma^2}{2\omega} & \text{otherwise} \end{cases}$$

- $\gamma \in \mathbb{R}$: Lagrange factor
- $\omega > 0$: penalty factor

Property of the augmented Lagrangian

If $\mathbf{x}^* \in \mathbb{R}^n$ satisfies KKT conditions, then for all $\omega > 0$

$$\nabla_{\mathbf{x}} h(\mathbf{x}^*,\gamma^*,\omega) = \nabla_{\mathbf{x}} f(\mathbf{x}^*) + \max(0, \gamma^* + \omega g(\mathbf{x}^*))\nabla_{\mathbf{x}} g(\mathbf{x}^*) = \mathbf{0}$$

New (unconstrained) optimization problem

$$\min_{\mathbf{x}} h(\mathbf{x},\gamma,\omega)$$

- Adaptive augmented Lagrangian approach: γ and ω are updated

Outline

Stochastic derivative-free optimization algorithm

$$\min_{\mathbf{x}} f(\mathbf{x}), f : \mathbb{R}^n \rightarrow \mathbb{R}, n \in \mathbb{N}$$

- Sequence $(\mathbf{s}_t)_{t \in \mathbb{N}}$ of states $\mathbf{s}_t$
- Transition function $\mathcal{F}^f$

$$\mathbf{s}_{t+1} = \mathcal{F}^f(\mathbf{s}_t, \mathbf{U}_{t+1})$$

  $\mathbf{U}_{t+1} = [\mathbf{U}^1_{t+1}, \cdots, \mathbf{U}^\lambda_{t+1}]$, $\mathbf{U}^k_{t+1}$ i.i.d. random vectors
- $\mathbf{s}_t$ includes a vector $\mathbf{X}_t \in \mathbb{R}^n$ representing the favorite solution at iteration t
- From $\mathbf{s}_t$ and $\mathbf{U}_{t+1}$, sample λ candidate solutions $\mathbf{X}^k_{t+1}$
- $\mathbf{X}^k_{t+1}$ evaluated on f

$$\min_{\mathbf{x}} f(\mathbf{x}) \text{ s.t. } g(\mathbf{x}) \leq 0 \quad \longrightarrow \quad \underbrace{h(\mathbf{x}, \gamma, \omega)}_{\text{augmented Lagrangian}} = f(\mathbf{x}) + \begin{cases} \gamma g(\mathbf{x}) + \frac{\omega}{2} g^2(\mathbf{x}) & \text{if } \gamma + \omega g(\mathbf{x}) \geq 0 \\ -\frac{\gamma^2}{2\omega} & \text{otherwise} \end{cases}$$

- Candidate solutions $\mathbf{X}_{t+1}^k$ evaluated on objective function

$$h_{(\gamma_t, \omega_t)}(\mathbf{X}_{t+1}^k) := h(\mathbf{X}_{t+1}^k, \gamma_t, \omega_t)$$

- State $\mathbf{s}_t{}' = [\mathbf{s}_t, \gamma_t, \omega_t]$ $\quad \rightarrow$ two additional state variables
- $\mathbf{s}_t{}'$ updated in two steps
  - $\mathbf{s}_{t+1} = \mathcal{F}^{h(\gamma_t, \omega_t)}(\mathbf{s}_t, \mathbf{U}_{t+1})$
  - Update γ_t, ω_t

- Update of Lagrange factor

$$\gamma_{t+1} = \max(0, \gamma_t + \omega_t g(\mathbf{X}_{t+1}))$$

- Update of penalty factor [Arnold et al.'15]

$$\omega_{t+1} = \begin{cases} \omega_t \chi^{1/4} & \text{if } \omega_t g^2(\mathbf{X}_{t+1}) < k_1 \frac{|h(\mathbf{X}_{t+1}, \gamma_t, \omega_t) - h(\mathbf{X}_t, \gamma_t, \omega_t)|}{n} \\ & \text{or } k_2 |g(\mathbf{X}_{t+1}) - g(\mathbf{X}_t)| < |g(\mathbf{X}_t)| \\ \mu_t \chi^{-1} & \text{otherwise} \end{cases}$$

→ avoid stagnation

→ avoid ill-conditioning

where

- $\omega_t g^2(\mathbf{X}_{t+1}) \approx |h(\mathbf{X}_{t+1}, \gamma_t + \Delta\lambda, \omega_t + \Delta\mu) - h(\mathbf{X}_{t+1}, \gamma_t, \omega_t)|$
- $\chi, k_1, k_2 \in \mathbb{R}^+_>$

Outline

Algorithm 1 $(\mu/\mu_w, \lambda)$-MSR-CMA-ES with Augmented Lagrangian Constraint Handling

given $n \in \mathbb{N}_{>}$, $\chi = 2^{1/n}$, $k_1 = 3$, $k_2 = 5$, $\mu, \lambda \in \mathbb{N}_{>}$, $j = 0.3\lambda$, $0 \leq w_i < 1$, $\sum_{i=1}^{\mu} w_i = 1$,

$$\mu_{\text{eff}} = 1/\sum_{i=1}^{\mu} w_i^2,\ c_\sigma = 0.3,\ d_\sigma = 2 - 2/n,\ c_c = \frac{4 + \mu_{\text{eff}}/n}{n + 4 - 2\mu_{\text{eff}}/n}$$

$$c_1 = \frac{2}{(n+1.3)^2 + \mu_{\text{eff}}},\ c_\mu = \min\left(1 - c_1, 2\frac{\mu_{\text{eff}} - 2 + 1/\mu_{\text{eff}}}{(n+2)^2 + \mu_{\text{eff}}}\right)$$

initialize $\mathbf{X}_0 \in \mathbb{R}^n$, $\sigma_0 \in \mathbb{R}^+_{>}$, $\mathbf{C}_0 = \mathbf{I}_{n \times n}$, $t = 0$, $q_0 = 0$, $p_0 = 0$,
    constrained_problem   // **true** if the problem is constrained, **false** otherwise

if constrained_problem

    initialize $\gamma_0 \in \mathbb{R}$, $\omega_0 \in \mathbb{R}^+_{>}$

while not happy

    $\mathbf{X}^i_{t+1} = \mathbf{X}_t + \sigma_t \mathbf{U}^i_{t+1}$, $\mathbf{U}^i_{t+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_t)$, $i = 1, \cdots, \lambda$ // sample candidate solutions

    Extract indices $\{1:\lambda, \cdots, \lambda:\lambda\}$ of ordered candidate solutions such that

$$\begin{cases} h(\mathbf{X}^{1:\lambda}_{t+1}, \gamma_t, \omega_t) \leq \ldots \leq h(\mathbf{X}^{\lambda:\lambda}_{t+1}, \gamma_t, \omega_t) & \text{if constrained\_problem} \\ f(\mathbf{X}^{1:\lambda}_{t+1}) \leq \ldots \leq f(\mathbf{X}^{\lambda:\lambda}_{t+1}) & \text{otherwise} \end{cases}$$

    $\mathbf{X}_{t+1} = \sum_{i=1}^{\mu} w_i \mathbf{X}^{i:\lambda}_{t+1} = \mathbf{X}_t + \sigma_t \sum_{i=1}^{\mu} w_i \mathbf{U}^{i:\lambda}_{t+1}$ // recombine μ best candidate solutions

$$K_{\text{succ}} = \begin{cases} \sum_{i=1}^{\lambda} \mathbf{1}_{\{h(\mathbf{X}^i_{t+1}, \gamma_t, \omega_t) \leq h(\mathbf{X}^{j:\lambda}_t, \gamma_t, \omega_t)\}} & \text{if constrained\_problem} \\ \sum_{i=1}^{\lambda} \mathbf{1}_{\{f(\mathbf{X}^i_{t+1}) \leq f(\mathbf{X}^{j:\lambda}_t)\}} & \text{otherwise} \end{cases}$$

    $z_t = \frac{2}{\lambda}\left(K_{\text{succ}} - \frac{\lambda}{2}\right)$ // compute success measure

    $q_{t+1} = (1 - c_\sigma) q_t + c_\sigma z_t$

    $\sigma_{t+1} = \sigma_t \exp\left(\frac{q_{t+1}}{d_\sigma}\right)$ // update step-size

    $p_{t+1} = (1 - c_c) p_t + \sqrt{c_c(2 - c_c)\mu_{\text{eff}}}\left(\frac{\mathbf{X}_{t+1} - \mathbf{X}_t}{\sigma_t}\right)$ // cumulation path for CMA

    $\mathbf{C}_{t+1} = (1 - c_1 - c_\mu)\mathbf{C}_t + c_1 p_{t+1} p_{t+1}^{\mathrm{T}} + c_\mu \sum_{i=1}^{\mu} w_i \left(\frac{\mathbf{X}^i_{t+1} - \mathbf{X}_t}{\sigma_t}\right)\left(\frac{\mathbf{X}^i_{t+1} - \mathbf{X}_t}{\sigma_t}\right)^{\mathrm{T}}$
    // update covariance matrix

    if constrained_problem

        $\gamma_{t+1} = \max(0, \gamma_t + \omega_t g(\mathbf{X}_{t+1}))$ // update Lagrange factor

$$\omega_{t+1} = \begin{cases} \omega_t \chi^{1/4} & \text{if } \omega_t g^2(\mathbf{X}_{t+1}) < k_1 \frac{|h(\mathbf{X}_{t+1}, \gamma_t, \omega_t) - h(\mathbf{X}_t, \gamma_t, \omega_t)|}{n} \\ & \text{or } k_2 |g(\mathbf{X}_{t+1}) - g(\mathbf{X}_t)| < |g(\mathbf{X}_t)| \\ \omega_t \chi^{-1} & \text{otherwise} \end{cases}$$
    // update penalty factor

    $\mathbf{X}^{j:\lambda}_{t+1} = \mathbf{X}_t + \sigma_t \mathbf{U}^{j:\lambda}_{t+1}$ // update jth best solution

    $t = t + 1$

Median success rule (MSR)

$$K_{succ} = \begin{cases} \sum_{i=1}^{\lambda} \mathbf{1}_{\{h(\mathbf{X}^i_{t+1}, \gamma_t, \omega_t) \le h(\mathbf{X}_t^{j:\lambda}, \gamma_t, \omega_t)\}} & \text{if constrained problem} \\ \sum_{i=1}^{\lambda} \mathbf{1}_{\{f(\mathbf{X}^i_{t+1}) \le f(\mathbf{X}_t^{j:\lambda})\}} & \text{otherwise} \end{cases}$$

$$z_t = \frac{2}{\lambda}\left(K_{succ} - \frac{\lambda}{2}\right) \quad \text{// compute success measure}$$

$$q_{t+1} = (1 - c_\sigma) q_t + c_\sigma z_t$$

$$\sigma_{t+1} = \sigma_t \exp\left(\frac{q_{t+1}}{d_\sigma}\right) \quad \text{// update step-size}$$

Covariance matrix adaptation ES (CMA-ES)

$$p_{t+1} = (1 - c_c) p_t + \sqrt{c_c(2 - c_c)\mu_{eff}} \left(\frac{\mathbf{X}_{t+1} - \mathbf{X}_t}{\sigma_t}\right) \quad \text{// cumulation path for CMA}$$

$$\mathbf{C}_{t+1} = (1 - c_1 - c_\mu)\mathbf{C}_t + c_1 p_{t+1} p_{t+1}^T + c_\mu \sum_{i=1}^{\mu} w_i \left(\frac{\mathbf{X}^i_{t+1} - \mathbf{X}_t}{\sigma_t}\right)\left(\frac{\mathbf{X}^i_{t+1} - \mathbf{X}_t}{\sigma_t}\right)^T$$

// update covariance matrix

Algorithm 1 $(\mu/\mu_w, \lambda)$-MSR-CMA-ES with Augmented Lagrangian Constraint Handling

given $n \in \mathbb{N}_+$, $\chi = 2^{1/n}$, $k_1 = 3$, $k_2 = 5$, $\mu, \lambda \in \mathbb{N}_+$, $j = 0.3\lambda$, $0 \leq w_i < 1$, $\sum_{i=1}^{\mu} w_i = 1$,

$$\mu_{\text{eff}} = 1/\sum_{i=1}^{\mu} w_i^2,\ c_\sigma = 0.3,\ d_\sigma = 2 - 2/n,\ c_c = \frac{4 + \mu_{\text{eff}}/n}{n + 4 - 2\mu_{\text{eff}}/n}$$

$$c_1 = \frac{2}{(n+1.3)^2 + \mu_{\text{eff}}},\ c_\mu = \min\left(1 - c_1, 2\frac{\mu_{\text{eff}} - 2 + 1/\mu_{\text{eff}}}{(n+2)^2 + \mu_{\text{eff}}}\right)$$

initialize $\mathbf{X}_0 \in \mathbb{R}^n$, $\sigma_0 \in \mathbb{R}_+^*$, $\mathbf{C}_0 = \mathbf{I}_{n\times n}$, $t = 0$, $q_0 = 0$, $p_0 = 0$,
constrained_problem // **true** if the problem is constrained, **false** otherwise

if constrained_problem

initialize $\gamma_0 \in \mathbb{R}$, $\omega_0 \in \mathbb{R}_+^*$

while not happy

$\mathbf{X}_{t+1}^i = \mathbf{X}_t + \sigma_t \mathbf{U}_{t+1}^i$, $\mathbf{U}_{t+1}^i \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_t)$, $i = 1, \cdots, \lambda$ // sample candidate solutions

Extract indices $\{1:\lambda, \cdots, \lambda:\lambda\}$ of ordered candidate solutions such that

$$\begin{cases} h(\mathbf{X}_{t+1}^{1:\lambda}, \gamma_t, \omega_t) \leq \ldots \leq h(\mathbf{X}_{t+1}^{\lambda:\lambda}, \gamma_t, \omega_t) & \text{if constrained\_problem} \\ f(\mathbf{X}_{t+1}^{1:\lambda}) \leq \ldots \leq f(\mathbf{X}_{t+1}^{\lambda:\lambda}) & \text{otherwise} \end{cases}$$

$\mathbf{X}_{t+1} = \sum_{i=1}^{\mu} w_i \mathbf{X}_{t+1}^{i:\lambda} = \mathbf{X}_t + \sigma_t \sum_{i=1}^{\mu} w_i \mathbf{U}_{t+1}^{i:\lambda}$ // recombine μ best candidate solutions

$$K_{succ} = \begin{cases} \sum_{i=1}^{\lambda} \mathbf{1}_{\{h(\mathbf{X}_{t+1}^{i}, \gamma_t, \omega_t) \leq h(\mathbf{X}_t^{j:\lambda}, \gamma_t, \omega_t)\}} & \text{if constrained\_problem} \\ \sum_{i=1}^{\lambda} \mathbf{1}_{\{f(\mathbf{X}_{t+1}^{i}) \leq f(\mathbf{X}_t^{j:\lambda})\}} & \text{otherwise} \end{cases}$$

$z_t = \frac{2}{\lambda}\left(K_{succ} - \frac{\lambda}{2}\right)$ // compute success measure

$q_{t+1} = (1 - c_\sigma) q_t + c_\sigma z_t$

$\sigma_{t+1} = \sigma_t \exp\left(\frac{q_{t+1}}{d_\sigma}\right)$ // update step-size

$p_{t+1} = (1 - c_c) p_t + \sqrt{c_c(2 - c_c)\mu_{\text{eff}}}\left(\frac{\mathbf{X}_{t+1} - \mathbf{X}_t}{\sigma_t}\right)$ // cumulation path for CMA

$\mathbf{C}_{t+1} = (1 - c_1 - c_\mu)\mathbf{C}_t + c_1 p_{t+1} p_{t+1}^{\mathsf{T}} + c_\mu \sum_{i=1}^{\mu} w_i \left(\frac{\mathbf{X}_{t+1}^{i} - \mathbf{X}_t}{\sigma_t}\right)\left(\frac{\mathbf{X}_{t+1}^{i} - \mathbf{X}_t}{\sigma_t}\right)^{\mathsf{T}}$
// update covariance matrix

if constrained_problem

$\gamma_{t+1} = \max(0, \gamma_t + \omega_t g(\mathbf{X}_{t+1}))$ // update Lagrange factor

$$\omega_{t+1} = \begin{cases} \omega_t \chi^{1/4} & \text{if } \omega_t g^2(\mathbf{X}_{t+1}) < k_1 \frac{|h(\mathbf{X}_{t+1}, \gamma_t, \omega_t) - h(\mathbf{X}_t, \gamma_t, \omega_t)|}{n} \\ & \text{or } k_2 |g(\mathbf{X}_{t+1}) - g(\mathbf{X}_t)| < |g(\mathbf{X}_t)| \\ \omega_t \chi^{-1} & \text{otherwise} \end{cases}$$ // update penalty factor

$\mathbf{X}_{t+1}^{j:\lambda} = \mathbf{X}_t + \sigma_t \mathbf{U}_{t+1}^{j:\lambda}$ // update jth best solution

$t = t + 1$

Outline

Experimental setting

- $f^{\alpha}_{\text{sphere}}(\mathbf{x}) = (\frac{1}{2}\sum_{i=1}^{n}\mathbf{x}_i^2)^{\alpha}$, $\alpha = 1, 2, 0.5$
- $f_{\text{ellipsoid}}(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{n}\alpha^{\frac{i-1}{n-1}}\mathbf{x}_i^2$, $\alpha = 10^2, 10^6$
- $f_{\text{diff\_pow}}(\mathbf{x}) = \sqrt{\sum_{i=1}^{n}|\mathbf{x}_i|^{2+4\frac{i-1}{n-1}}}$
- Unimodal problems: $\mathbf{x}_{\text{opt}} = (10, \cdots, 10)^{\mathsf{T}}$, $\gamma_{\text{opt}} = 1$
- $g(\mathbf{x}) = \mathrm{b}^{\mathsf{T}}\mathbf{x} + c$, $\mathrm{b} = -\nabla_{\mathrm{x}} f_{.}(\mathbf{x}_{\text{opt}})^{\mathsf{T}}$, $c = \nabla_{\mathrm{x}} f_{.}(\mathbf{x}_{\text{opt}})\mathbf{x}_{\text{opt}}$
- $n = 10, 100$
- $\mathbf{X}_0$ sampled uniformly in $[-5, 5]^n$, $\sigma_0 = 1$, $\gamma_0 = 5$, $\omega_0 = 1$
- $\lambda =$, $\mu = \lambda/2$

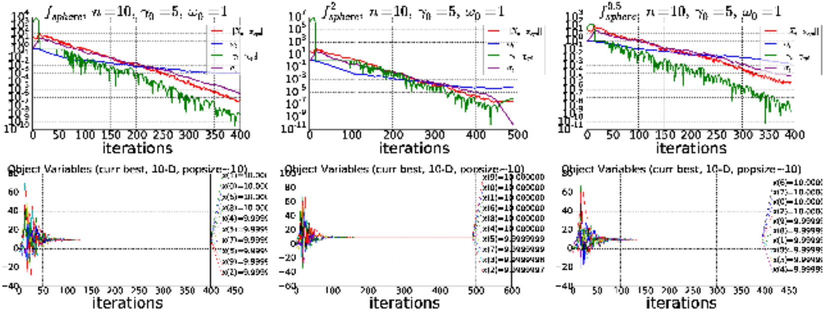


Figure: Single runs of $(\mu/\mu_w, \lambda)$-MSR-CMA-ES with augmented Lagrangian on $f_{\mathbf{sphere}}$ (left), $f^2_{\mathbf{sphere}}$ (middle), and $f^{\mathbf{0.5}}_{\mathbf{sphere}}$ (right) in $n = 10$. The optimum $\mathbf{x_{opt}} = (10, \cdots, 10)^\mathsf{T}$. Top: evolution of the distance to the optimum, the distance to the Lagrange multiplier, the penalty factor, and the step-size in log-scale. Bottom: evolution of the coordinates of $\mathbf{X}_t$

Empirical Results

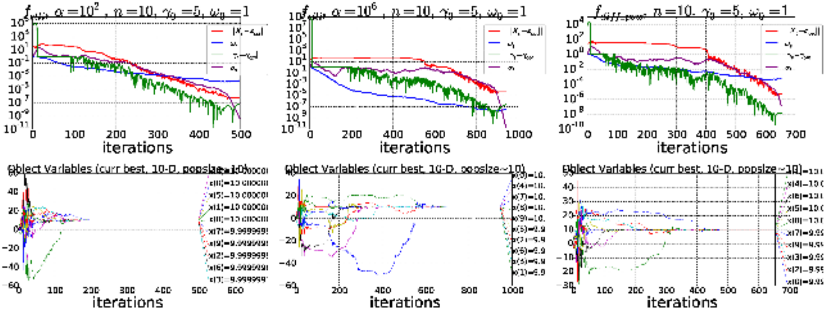


Figure: Single runs of $(\mu/\mu_w, \lambda)$-MSR-CMA-ES with augmented Lagrangian on $f_{\text{ellipsoid}}$ with $\alpha = 10^2$ (left), $f_{\text{ellipsoid}}$ with $\alpha = 10^6$ (middle), and $f_{\text{diff\_pow}}$ (right) in $n = 10$. The optimum $\mathbf{x}_{\text{opt}} = (10, \cdots, 10)^{\mathsf{T}}$. Top: evolution of the distance to the optimum, the distance to the Lagrange multiplier, the penalty factor, and the step-size in log-scale. Bottom: evolution of the coordinates of $\mathbf{X}_t$

Outline

Definition: Linear convergence

Consider an ES minimizing a function $f : \mathbb{R}^n \to \mathbb{R}$. The sequence $(\mathbf{X}_t)_{t\in\mathbb{N}}$ of the solutions computed by the algorithm converges linearly to the minimum x_{opt} of f if

$$\lim_{t\to\infty} \frac{1}{t} \ln \frac{\|\mathbf{X}_t - x_{\text{opt}}\|}{\|\mathbf{X}_0 - x_{\text{opt}}\|} = -\text{CR} \quad a.s.$$

where CR > 0 is the convergence rate

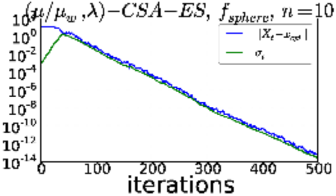


Figure: Single run of $(\mu/\mu_w,\lambda)$-CSA-ES on unconstrained f_{sphere}

- $(\mathbf{s}_t)_{t\in\mathbb{N}}$: sequence of states of a stochastic DFO algorithm minimizing f
- Construct homogeneous Markov chain from state variables
- Prove its stability: φ-irreducibility, positivity, Harris-recurrence
- Express convergence rate as a function of the Markov chain
- Apply law of large numbers for Markov chains to deduce linear convergence/divergence

- $(\mu/\mu_w, \lambda)$-MSR-ES (without CMA)
- $h(\mathbf{x}, \gamma, \omega) = f(\mathbf{x}) + \gamma g(\mathbf{x}) + \frac{\omega}{2} g^2(\mathbf{x})$
- $\gamma_{t+1} = \gamma_t + \omega_t g(\mathbf{X}_{t+1})$
- If h positive homogeneous of degree 2 w.r.t. $(\bar{\mathbf{x}}, \bar{\gamma})$ where $\bar{\mathbf{x}} \in \mathbb{R}^n$, $\bar{\gamma} \in \mathbb{R}$, $g(\bar{\mathbf{x}}) = 0$, then $(\mathbf{Y}_t, q_t, R_t, \Gamma_t, \omega_t)_{t\in\mathbb{N}}$ is homogeneous Markov chain

$$\mathbf{Y}_t = \frac{\mathbf{X}_t - \bar{\mathbf{x}}}{\sigma_t},\ R_t = \frac{\mathbf{X}_t^{j:\lambda} - \bar{\mathbf{x}}}{\sigma_t},\ \Gamma_t = \frac{\gamma_t - \bar{\gamma}}{\sigma_t}$$

- If $(\mathbf{Y}_t, q_t, R_t, \Gamma_t, \omega_t)_{t\in\mathbb{N}}$ is φ-irreducible, positive, Harris-recurrent with invariant probability measure π

$$\lim_{t\to\infty} \frac{1}{t} \ln \frac{\|\mathbf{X}_t - \mathbf{x}_{\text{opt}}\|}{\|\mathbf{X}_0 - \mathbf{x}_{\text{opt}}\|} = \lim_{t\to\infty} \frac{1}{t} \ln \frac{|\gamma_t - \gamma_{\text{opt}}|}{|\gamma_0 - \gamma_{\text{opt}}|} = \lim_{t\to\infty} \frac{1}{t} \ln \frac{\sigma_t}{\sigma_0} = -\text{CR}$$

where

$$-\text{CR} = \underbrace{\text{E}_\pi(\mathcal{R}(\Phi))}_{\text{expectation of step-size change}}$$

- Stability checked empirically

Conclusion

- General framework for building stochastic DFO algorithm for constrained optimization with augmented Lagrangian constraint handling
- Linear convergence of $(\mu/\mu_w, \lambda)$-MSR-CMA-ES with adaptive augmented Lagrangian on convex quadratic and ill-conditioned functions with one linear inequality constraint
  - "Simple" constrained problem (one inequality constraint)
  - Necessary to understand whether it is possible to converge linearly [Atamna et al.'16]
- Extension to many constraints possible

→ design new update rules for γ and ω