Linear Convergence of Evolution Strategies with Derandomized Sampling

Beyond Quasi-Convex Functions

Jeremie Decock

Olivier Teytaud

Inria

May 28, 2014

Introduction

The aim of this presentation

Study convergence rate of a simple pattern search method

Introduction

Convergence of evolutionary algorithms

Proofs are almost always for (quasi-)convex objective functions

Non quasi-convex objective functions

- some proofs for convergence (asymptotically the optimum is found)
- few for linear convergence (precision $O\left(e^{-\Omega(n)}\right)$ after n iterations)
    - in discrete space
    - not in continuous space

Introduction

Our contribution

- Prove the linear convergence of an algorithm
  - on non quasi-convex functions
  - on continuous domains
- Under some assumptions about
  - the sampling performed by the algorithm
  - the “conditioning” of the objective function
  - the unicity of the optimum

Overview

Framework

Algorithm

Initialize $\mathbf{x} \in \mathbb{R}^d$
Parameters $k \in \mathbb{N}^*, \boldsymbol{\delta}_1, \dots, \boldsymbol{\delta}_k \in \mathbb{R}^d, \sigma \in \mathbb{R}_+^*, k_1 \in \mathbb{N}^*, k_2 \in \mathbb{N}^*$
for $t = 1, 2, 3, \dots$ **do**

// **mutations**
For $i \in [[1, k]]$, $\mathbf{x}_i \leftarrow \mathbf{x} + \sigma \boldsymbol{\delta}_i$

// **useful auxiliary variables**
$n \leftarrow$ number of $\mathbf{x}_i$ such that $f(\mathbf{x}_i) < f(\mathbf{x})$
$\mathbf{x}' \leftarrow \mathbf{x}_i$ with $i \in [[1, k]]$ such that $f(\mathbf{x}_i)$ is minimum

...

end for

Algorithm

Initialize $\mathbf{x} \in \mathbb{R}^d$
Parameters $k \in \mathbb{N}^*, \boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_k \in \mathbb{R}^d, \sigma \in \mathbb{R}_+^*, k_1 \in \mathbb{N}^*, k_2 \in \mathbb{N}^*$
for $t = 1, 2, 3, \ldots$ **do**

...
// **step-size adaptation**
if $n \leq k_1$ **then**
$\sigma \leftarrow \sigma/2$
end if
if $n \geq k_2$ **then**
$\sigma \leftarrow 2\sigma$
end if
...

end for

Algorithm

Initialize $\mathbf{x} \in \mathbb{R}^d$
Parameters $k \in \mathbb{N}^*, \boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_k \in \mathbb{R}^d, \sigma \in \mathbb{R}_+^*, k_1 \in \mathbb{N}^*, k_2 \in \mathbb{N}^*$
for $t = 1, 2, 3, \ldots$ **do**

...

// **win: accepted mutation**
if $k_1 < n < k_2$ **then**
$\mathbf{x} \leftarrow \mathbf{x}'$
end if

end for

Assumptions

Objective function

The objective function f is unimodal

The considered algorithms are invariant by translation or composition with increasing functions, therefore we can state that

- $\mathbf{x}^* = \mathbf{0}$ is the optimum
- $f(\mathbf{x}^*) = 0$

Assumptions

Conditioning of f

Conditioning of f: $\exists K' > 0,\ \exists K'' > 0$ s.t. $\forall \mathbf{x} \in \mathbb{R}^d$

$$K'||\mathbf{x}|| \leq f(\mathbf{x}) \leq K''||\mathbf{x}||$$

This assumption is not so strong as a constraint and in fact, quadratic positive definite forms with bounded condition number are covered

Assumptions

Deterministic sampling of the algorithm

The sampling of the algorithm is deterministic (like in pattern search methods)

- mutation vectors $\boldsymbol{\delta}_i$ are constant
- the evolution of the step size parameter σ is deterministic

Assumptions

Regular sampling of the algorithm

We assume $\exists b, b', c', c, \eta$ s.t.
$0 < b < b' \leq 2b' \leq c' \leq c,\ 0 < \eta < 1,\ \forall \mathbf{x} \in \mathbb{R}^d$

$$\sigma \geq b^{-1}||\mathbf{x}|| \quad \Rightarrow \quad n \leq k_1 \qquad (\sigma \text{ too large}) \qquad (1)$$

$$\sigma \leq b'^{-1}||\mathbf{x}|| \quad \Rightarrow \quad n > k_1 \qquad (\sigma \text{ small enough}) \qquad (2)$$

$$\sigma \geq c'^{-1}||\mathbf{x}|| \quad \Rightarrow \quad n < k_2 \qquad (\sigma \text{ large enough}) \qquad (3)$$

$$\sigma \leq c^{-1}||\mathbf{x}|| \quad \Rightarrow \quad n \geq k_2 \qquad (\sigma \text{ too small}) \qquad (4)$$

$$b'^{-1}||\mathbf{x}|| \leq \sigma \leq c'^{-1}||\mathbf{x}|| \quad \Rightarrow \quad \exists i \in [[1,k]]; f(\mathbf{x}_i) \leq \eta f(\mathbf{x}) \qquad (5)$$

with $n := \#\{i \in [[1,k]]; f(\mathbf{x} + \sigma\boldsymbol{\delta}_i) < f(\mathbf{x})\}$

Theorem

Preliminary work

Define $l = \ln\left(\frac{||\mathbf{x}||}{\sigma}\right)$. Eqs. 1-5 can be rephrased as follows:

$$l \leq \ln(b) \Rightarrow n \leq k_1 \quad (\sigma \textit{ too large}) \tag{6}$$

$$l \geq \ln(b') \Rightarrow n > k_1 \quad (\sigma \textit{ small enough}) \tag{7}$$

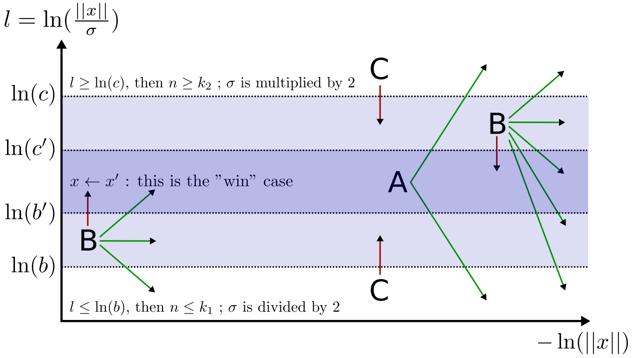
$$l \leq \ln(c') \Rightarrow n < k_2 \quad (\sigma \textit{ large enough}) \tag{8}$$

$$l \geq \ln(c) \Rightarrow n \geq k_2 \quad (\sigma \textit{ too small}) \tag{9}$$

$$\ln(b') \leq l \leq \ln(c') \Rightarrow \exists i \in [[1,k]]; f(\mathbf{x} + \sigma\boldsymbol{\delta}_i) \leq \eta f(\mathbf{x}) \tag{10}$$

with $n := \#\{i \in [[1,k]]; f(\mathbf{x} + \sigma\boldsymbol{\delta}_i) < f(\mathbf{x})\}$

Preliminary work



Preliminary work

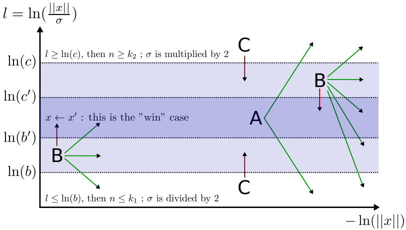
Forced increase

Forced increase

if $l \leq \ln(b)$, then

- $n \leq k_1$
- σ is divided by 2
- l is increased by $\ln(2)$ (Eq. 6)

This is a case C at the bottom in the figure



Preliminary work

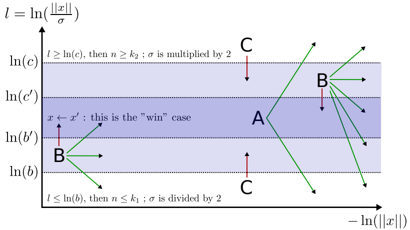
Forced decrease

Forced decrease

if $l \geq \ln(c)$, then

- $n \geq k_2$
- σ is multiplied by 2
- l is decreased by $\ln(2)$ (Eq. 9)

This is a case C at the top in the figure



Preliminary work

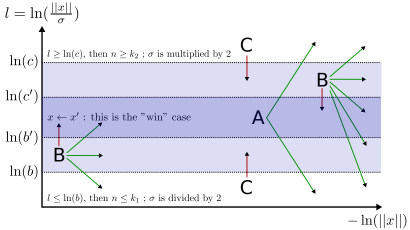
Forced win

Forced win

if $\ln(b') \leq l \leq \ln(c')$, then

- this is the "sure win" case (Eq. 10)
- $\mathbf{x} \leftarrow \mathbf{x}'$ ($\mathbf{x}'$ is the best $\mathbf{x}_i$)
- l can be
  - increased (at most by $\max_i ||\boldsymbol{\delta}_i||$) or
  - decreased (by $\Delta = \ln\left(\frac{||\mathbf{x}||}{||\mathbf{x}'||}\right)$)

This is a case A in the figure



Preliminary work

Uncertain outcome

Uncertain outcome

if $\ln(b) \leq l \leq \ln(b')$ or $\ln(c) \leq l \leq \ln(c')$, then

- the iteration can be an improvement or not
- if not, the point is moved towards case A with steps of $\ln(2)$
- if there's no "win" case, then in the mean time l will arrive between $\ln(b')$ and $\ln(c')$, where a win is ensured

This is a case B in the previous figure

Theorem

There exists a constant K, depending on $\eta, K', K'', \max_i ||\boldsymbol{\delta}_i||$ only such that for index t large enough

$$\frac{\ln(||\mathbf{X}_t||)}{t} \leq K < 0$$

where $\mathbf{X}_t$ is the tested solution $\mathbf{x}$ at iteration t

Proof

Step 1

Showing that there are infinitely many wins

1. l is increased or decreased when it is too low or too high
   - the algorithm eventually brings l to the "win" range
2. l can be increased or decreased at most by $\ln(2)$ and $b' \leq 2b' \leq c'$
   - the algorithm can not jump over the "win" range

This ensures that infinitely often we have a "win" : $\mathbf{x} \leftarrow \mathbf{x}'$

Proof

Step 2

Showing that “wins” are big enough

“Win” case implies

- $f(\mathbf{x}') \leq \eta f(\mathbf{x})$
- $f(\mathbf{x}') \leq K''||\mathbf{x}'|| \leq \frac{K''}{K'}\frac{||\mathbf{x}'||}{||\mathbf{x}||}f(\mathbf{x})$

so that $\ln(f(\mathbf{x}))$ is decreased by at least

$$\max\left(\ln\left(\frac{1}{\eta}\right)\ ,\ \ln\left(\frac{K'}{K''}\right)+\ln\left(\frac{||\mathbf{x}||}{||\mathbf{x}'||}\right)\right) \tag{11}$$

Proof

Step 2

Showing that the number of steps between two "wins" is low enough

After a "win", the number of iterations to the next "win" is

- at most $z = 1 + \ln(\frac{c}{b})\frac{\Delta}{\ln(2)}$ if $l' \leq \ln(b')$
- at most $z = 1 + \frac{\max_i ||\boldsymbol{\delta}_i||}{\ln(2)}$ if $l' \geq \ln(c')$
- less than both cases above otherwise

with $l' = \ln\left(\frac{||\mathbf{x}'||}{\sigma}\right)$ and $\Delta = \ln\left(\frac{||\mathbf{x}||}{||\mathbf{x}'||}\right)$

Proof

Step 2

Progress rate

Eq. 11 divided by z is lower bounded by some positive constant

$$\begin{aligned} \text{ProgressRate} &= \text{Eq. 11 divided by } z \\ &= \frac{\max\left(\ln\left(\frac{1}{\eta}\right)\ ,\ \ln\left(\frac{K'}{K''}\right)+\Delta\right)}{\min\left(1+\ln\left(\frac{c}{b}\right)\frac{\Delta}{\ln(2)}\ ,\ 1+\frac{\max_j |||\delta_j||}{\ln(2)}\right)} \end{aligned}$$

Proof

Step 3

Summing iterations

Hence if $t > n_0$,

$$\ln(f(\mathbf{X}_t)) \leq \ln(f(\mathbf{X}_1)) - (t - n_0) \times \sum_i \frac{\max\left(\ln\left(\frac{1}{\eta}\right),\ \ln\left(\frac{K'}{K''}\right) + \Delta\right)}{\min\left(1 + \ln(\frac{c}{b})\frac{\Delta}{\ln(2)}\ ,\ 1 + \frac{\max_j ||\boldsymbol{\delta}_j||}{\ln(2)}\right)} \quad \square$$

where :

- summation is for i index of an iteration t with a "win"
- n_0 is the number of initial iterations before a "win"

Application to quadratic functions

Application to quadratic functions

Considered family of objective functions

f is quadratic positive definite objective functions such that

$$\frac{maxEigenValue(Hessian(f))}{minEigenValue(Hessian(f))} < c_{max} < \infty$$

Consider Q a positive definite quadratic form with optimum in 0

We work on $\mathbf{x} \mapsto \sqrt{Q(\mathbf{x}-\mathbf{x}^*)}$ instead of $\mathbf{x} \mapsto Q(\mathbf{x}-\mathbf{x}^*)$ so that the first assumption is verified:

$$K'||\mathbf{x}|| \leq f(\mathbf{x}) \leq K''||\mathbf{x}||$$

$\forall \mathbf{x} \in \mathbb{R}^d, \ \exists K' > 0, \ \exists K'' > 0$

Application to quadratic functions

Assumptions to verify (reminder)

We try to prove that f respects the following assumptions

$\exists b, b', c', c, \eta$ s.t. $0 < b < b' \leq 2b' \leq c' \leq c,\ 0 < \eta < 1,\ \forall \mathbf{x} \in \mathbb{R}^d$

$$
\begin{aligned}
\sigma \geq b^{-1}||\mathbf{x}|| \quad &\Rightarrow \quad n \leq k_1 \qquad (\sigma \text{ too large}) \\
\sigma \leq b'^{-1}||\mathbf{x}|| \quad &\Rightarrow \quad n > k_1 \qquad (\sigma \text{ small enough}) \\
\sigma \geq c'^{-1}||\mathbf{x}|| \quad &\Rightarrow \quad n < k_2 \qquad (\sigma \text{ large enough}) \\
\sigma \leq c^{-1}||\mathbf{x}|| \quad &\Rightarrow \quad n \geq k_2 \qquad (\sigma \text{ too small}) \\
b'^{-1}||\mathbf{x}|| \leq \sigma \leq c'^{-1}||\mathbf{x}|| \quad &\Rightarrow \quad \exists i \in [[1, k]]; f(\mathbf{x}_i) \leq \eta f(\mathbf{x})
\end{aligned}
$$

with $n := \#\{i \in [[1, k]]; f(\mathbf{x} + \sigma \boldsymbol{\delta}_i) < f(\mathbf{x})\}$

Application to quadratic functions

We note:

- $p = p_{\mathbf{x},\sigma,f}$ the probability that $\mathbf{x} + \sigma\boldsymbol{\delta}_i$ is in $E = f^{-1}([0, f(\mathbf{x})[)$
- $\hat{p} = \hat{p}_{\mathbf{x},\sigma,f}$ the frequency $\frac{1}{k}\sum_{i=1}^{k} \mathbf{1}_{\mathbf{x}+\sigma\boldsymbol{\delta}_i \in E}$

The previous assumptions essentially mean that frequencies are close to expectations for

- $\mathbf{x} + \sigma\boldsymbol{\delta}_i \in f^{-1}([0, f(\mathbf{x})[)$
- $\mathbf{x} + \sigma\boldsymbol{\delta}_i \in f^{-1}([0, \eta f(\mathbf{x})[)$

uniformly in $\mathbf{x}$, σ, f.

Corollary

Application of the main theorem to quadratic forms

Assume that the $\boldsymbol{\delta}_i$ are uniformly randomly drawn in the unit ball $B(0, 1)$.

Assume that F is the set of quadratic functions with minimum in 0 ($f(0) = 0$) as defined before.

Then, almost surely on the sequence $\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_k$, for k large enough and some parameters k_1 and k_2 of our evolutionary algorithm, then assumptions hold, and therefore for some $K < 0$, for all $t > 0$,

$$\frac{\ln(||\mathbf{X}_t||)}{t} \leq K$$

where $\mathbf{X}_t$ is the tested solution $\mathbf{x}$ at iteration t

Proof

Step 1

Using VC-dimension for approximating expectations by frequencies

The finiteness of the VC-dimension of quadratic forms state that for all $\epsilon > 0$, almost surely in $\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_k$, for all $\delta > 0$ and k sufficiently large, with probability at least $1 - \delta$,

$$\sup_{\mathbf{x}, f, \sigma > 0} \left| \hat{p}_{\mathbf{x},\sigma,f} - p_{\mathbf{x},\sigma,f} \right| \leq \epsilon/2$$

where $\mathbf{x}$ ranges over the domain, f ranges over F

Proof

Step 2

Showing that small σ leads to high acceptance rate and high σ leads to small acceptance rate

Thanks to the bounded conditioning, there exists $\epsilon > 0$ s.t.

$$s' < \frac{1}{2} s$$

$$\text{with} \qquad s = \sup\left\{\frac{\sigma}{||\mathbf{x}||}; \sigma, \mathbf{x}, f \text{ s.t. } p \geq \frac{\epsilon}{2}\right\}$$

$$\text{and} \quad s' = \inf\left\{\frac{\sigma}{||\mathbf{x}||}; \sigma, \mathbf{x}, f \text{ s.t. } p < \frac{1}{2} - \frac{\epsilon}{2}\right\}$$

Indeed $s' \to 0$ and $s \to \infty$ as $\epsilon \to 0$.

Proof

Step 2

Showing that small σ leads to high acceptance rate and high σ leads to small acceptance rate

The previous equations provide k_1, k_2, c' and b'

$$\begin{aligned}
\frac{1}{b'} &= \sup\left\{\frac{\sigma}{||\mathbf{x}||}; \sigma, \mathbf{x}, f \text{ s.t. } \hat{p} \geq \epsilon\right\} \\
\frac{1}{c'} &= \inf\left\{\frac{\sigma}{||\mathbf{x}||}; \sigma, \mathbf{x}, f \text{ s.t. } \hat{p} < \frac{1}{2} - \epsilon\right\} \\
k_1 &= \lfloor \epsilon k \rfloor \\
k_2 &= \left\lceil (\frac{1}{2} - \epsilon) k \right\rceil
\end{aligned}$$

Equations above imply $c' \geq 2b'$

Proof

Step 3

Showing that k large enough and σ well chosen leads to at least one mutation with significant improvement

Similarly, k large enough yield

$$
\begin{aligned}
b^{-1} &= \sup\left\{\frac{\sigma}{||\mathbf{x}||};\sigma,\mathbf{x},f \text{ s.t. } \hat{p} > k_1/k\right\} \\
c^{-1} &= \inf\left\{\frac{\sigma}{||\mathbf{x}||};\sigma,\mathbf{x},f \text{ s.t. } \hat{p} < k_2/k\right\}
\end{aligned}
$$

which provide Eqs. 4 and 1 with $b < c$.
Eqs. 1-4 then imply $b < b'$ and $c' < c$.

Proof

Step 3

Showing that k large enough and σ well chosen leads to at least one mutation with significant improvement

We now have to ensure the last assumption:

$$b'^{-1}||\mathbf{x}|| \leq \sigma \leq c'^{-1}||\mathbf{x}|| \Rightarrow \exists i \in [[1,k]]; f(\mathbf{x}+\sigma\boldsymbol{\delta}_i) \leq \eta f(\mathbf{x})$$

For now on, we note:

- $q = q_{\mathbf{x},\sigma,f}$ the probability that $\mathbf{x}+\sigma\boldsymbol{\delta}_i$ is in $E' = f^{-1}([0,\eta f(\mathbf{x})[)$
- $\hat{q} = \hat{q}_{\mathbf{x},\sigma,f}$ the frequency $\frac{1}{k}\sum_{i=1}^{k} \mathbf{1}_{\mathbf{x}+\sigma\boldsymbol{\delta}_i \in E'}$

Proof

Step 3

Showing that k large enough and σ well chosen leads to at least one mutation with significant improvement

Lets us assume

$$b^{-1} \leq \frac{\sigma}{||\mathbf{x}||} \leq c^{-1}$$

this implies $q > \epsilon_0$ for some $\epsilon_0 > 0$

For k sufficiently large for ensuring $\sup_{\sigma,\mathbf{x},f} |q_{\mathbf{x},\sigma,f} - \hat{q}_{\mathbf{x},\sigma,f}| \leq \epsilon_0/2$, by VC-dimension, we get $q' \geq \epsilon_0/2 > 0$

This implies that at least one $\boldsymbol{\delta}_i$ verifies $\mathbf{x} + \boldsymbol{\delta}_i \in E'$.
This is the last assumption.

Proof

Step 4

Concluding

We have shown our assumptions for square roots of positive definite quadratic normal forms with bounded conditioning. Therefore, the main theorem can be applied and leads to

$$\frac{\ln(||\mathbf{X}_t||)}{t} \leq K < 0$$

Conclusion

Conclusion

This is the first proof of linear convergence of an evolutionary algorithm in continuous domains on non quasi-convex functions.

Even the application to quadratic positive definite forms is new.

Future work

- Evaluate the convergence rate as a function of condition numbers
- Extend results to randomized algorithms

Thank you for your attention

Questions ?

Theorem's proof

Step 2

$$(11) \iff \ln\left(\frac{f(\mathbf{x})}{f(\mathbf{x}')}\right) \geq \max\left(\ln\left(\frac{1}{\eta}\right)\ ,\ \ln\left(\frac{K'}{K''}\right) + \ln\left(\frac{||\mathbf{x}||}{||\mathbf{x}'||}\right)\right)$$

$$\begin{aligned}
f(\mathbf{x}') \leq \eta f(\mathbf{x}) \quad & \iff \quad \frac{1}{\eta} f(\mathbf{x}') \leq f(\mathbf{x}) \\
& \iff \quad \frac{1}{\eta} \leq \frac{f(\mathbf{x})}{f(\mathbf{x}')} \\
& \iff \quad \ln\left(\frac{f(\mathbf{x})}{f(\mathbf{x}')}\right) \geq \ln\left(\frac{1}{\eta}\right) \\
f(\mathbf{x}') \leq \frac{K''}{K'} \frac{||\mathbf{x}'||}{||\mathbf{x}||} f(\mathbf{x}) \quad & \iff \quad f(\mathbf{x}') \frac{K'}{K''} \frac{||\mathbf{x}||}{||\mathbf{x}'||} \leq f(\mathbf{x}) \\
& \iff \quad \frac{K'}{K''} \frac{||\mathbf{x}||}{||\mathbf{x}'||} \leq \frac{f(\mathbf{x})}{f(\mathbf{x}')} \\
& \iff \quad \ln\left(\frac{K'}{K''} \frac{||\mathbf{x}||}{||\mathbf{x}'||}\right) \leq \ln\left(\frac{f(\mathbf{x})}{f(\mathbf{x}')}\right) \\
& \iff \quad \ln\left(\frac{f(\mathbf{x})}{f(\mathbf{x}')}\right) \geq \ln\left(\frac{K'}{K''}\right) + \ln\left(\frac{||\mathbf{x}||}{||\mathbf{x}'||}\right)
\end{aligned}$$

Theorem's proof

Step 2

if $l' \geq \ln(c')$ then

- $l < ln(c)$ (otherwise it couldn't be a "win")

$$
\begin{aligned}
l' - \ln(c') &\leq \ln(c + \max_i ||\delta_i||) - ln(c') \\
&\leq \ln(c) + \ln(1 + \max_i ||\delta_i||/c) - \ln(c') \\
&\leq \ln(c/c') + \max_i ||\delta_i||/c \\
&\leq \max_i ||\delta_i||/c
\end{aligned}
$$

$$
\begin{aligned}
l' - \ln(c') = \ln(\frac{||x + \sigma\delta_i||}{\sigma}) - \ln(c') &\leq \ln(\frac{||x||}{\sigma} + \delta_i) - \ln(c') \\
&\leq \ln(c * (1 + \frac{\delta_i}{c})) - \ln(c') \\
&\leq \ln(c) + \ln(1 + \max_i ||\delta_i||/c) - \ln(c') \\
&\leq \ln(c/c') + \max_i ||\delta_i||/c \\
&\leq \max_i ||\delta_i||/c
\end{aligned}
$$

Theorem's Proof

Step 3

Summing iterations

Hence if $t > n_0$,

$$\ln(f(\mathbf{X}_t)) \leq \ln(f(\mathbf{X}_1)) - (t - n_0) \times \sum_i \frac{\max\left(\ln\left(\frac{1}{\eta}\right),\ \ln\left(\frac{K'}{K''}\right) + \Delta\right)}{\min\left(1 + \ln(\frac{c}{b})\frac{\Delta}{\ln(2)}\ ,\ 1 + \ln(\frac{c}{b})\frac{\max_j \ln(||\boldsymbol{\delta}_j||)}{\ln(2)}\right)}$$

$$\iff \quad \ln(f(\mathbf{X}_t)) - \ln(f(\mathbf{X}_1)) \leq -(t - n_0) \times C$$

$$\iff \quad \frac{\ln\left(\frac{f(\mathbf{X}_t)}{f(\mathbf{X}_1)}\right)}{t - n_0} \leq -C$$

$$\Rightarrow \quad \frac{\ln(||\mathbf{X}_t||)}{t} \leq K < 0 \quad \text{(theorem)}$$

with C a positive constant.

Corollary's proof

Step 2 (1/4)

Step 2: showing that σ small leads to high acceptance rate and σ high leads to small acceptance rate.
Thanks to the bounded conditioning, there exists $\epsilon > 0$ s.t.

$$s' < \frac{1}{2}s \quad (12)$$

$$\text{with } s = \sup\left\{\frac{\sigma}{||\mathbf{x}||}; \sigma, \mathbf{x}, f \text{ s.t. } p \geq \frac{\epsilon}{2}\right\}$$

$$\text{and } s' = \inf\left\{\frac{\sigma}{||\mathbf{x}||}; \sigma, \mathbf{x}, f \text{ s.t. } p < \frac{1}{2} - \frac{\epsilon}{2}\right\}$$

because $s' \to 0$ and $s \to \infty$ as $\epsilon \to 0$.

Corollary's proof

Step 2 (2/4)

Notes

$$\hat{s} = \sup\left\{\frac{\sigma}{||\mathbf{x}||}; \sigma, \mathbf{x}, f \text{ s.t. } \hat{p} \geq \epsilon\right\}$$

$$\hat{s'} = \inf\left\{\frac{\sigma}{||\mathbf{x}||}; \sigma, \mathbf{x}, f \text{ s.t. } \hat{p} < \frac{1}{2} - \epsilon\right\}$$

Then

$$\sup_{\mathbf{x}, f, \sigma > 0} |\hat{p} - p| \leq \epsilon/2$$

implies

$$\frac{1}{2}\hat{s} \geq \frac{1}{2}s \quad \text{and} \quad s' \geq \hat{s'}$$

Corollary's proof

Step 2 (3/4)

So

$$\hat{s}' \leq s' < \frac{1}{2} s \leq \frac{1}{2} \hat{s}$$

and

$$\hat{s}' \leq \frac{1}{2} \hat{s}$$

Corollary's proof

Step 2 (4/4)

This provides k_1, k_2, c' and b' as follows for Eqs. 3 and 2:

$$\begin{aligned}
\frac{1}{b'} &= \hat{s} = \sup\left\{\frac{\sigma}{||\mathbf{x}||};\sigma,\mathbf{x},f \text{ s.t. } \hat{p}\geq\epsilon\right\} \\
\frac{1}{c'} &= \hat{s'} = \inf\left\{\frac{\sigma}{||\mathbf{x}||};\sigma,\mathbf{x},f \text{ s.t. } \hat{p}<\frac{1}{2}-\epsilon\right\} \\
k_1 &= \lfloor \epsilon k \rfloor \\
k_2 &= \left\lceil (\frac{1}{2}-\epsilon)k \right\rceil
\end{aligned}$$

Eqs. above imply $c' \geq 2b'$

Corollary's proof

Step 3

k large enough yield

$$\begin{aligned} b^{-1} &= \sup\left\{\frac{\sigma}{||\mathbf{x}||}; \sigma, \mathbf{x}, f \text{ s.t. } \hat{p} > k_1/k\right\}, \\ c^{-1} &= \inf\left\{\frac{\sigma}{||\mathbf{x}||}; \sigma, \mathbf{x}, f \text{ s.t. } \hat{p} < k_2/k\right\}, \end{aligned}$$

which provide assumptions 2 and 5 with $b < c$

Assumptions 2 and 5 then imply $b < b'$ and $c' < c$

Non quasi-convex functions