DYNAMIC PROGRAMMING OPTIMIZATION OVER RANDOM DATA: THE SCALING EXPONENT FOR NEAR-OPTIMAL SOLUTIONS

Marc Lelarge

INRIA-ENS

Joint work with **David Aldous** (UC Berkeley) and **Charles Bordenave** (Toulouse)

YEP - EURANDOM

"REGULAR" NEAR-MINIMAL PROBLEM

- Cost function $F : \mathbb{R}^n \longrightarrow \mathbb{R}$ reaching its minimum at x^*.

- Relation between the distance $\delta = |x - x^*|$ and the difference $\varepsilon = F(x) - F(x^*)$:

  if F is "smooth",

$$\varepsilon(\delta) = \inf\{F(x) - F(x^*),\ |x - x^*| \geq \delta\} \sim C\delta^2.$$

$\Longrightarrow$ the scaling exponent corresponding to this problem is 2.

NEAR-MINIMAL SOLUTIONS IN COMBINATORIAL OPTIMIZATION

[Aldous-Percus (2003)]

Consider a combinatorial optimization problem with n constituents under a suitable probability model.

As n goes to infinity, how cost increases when the optimal solution undergoes a small perturbation δ ?

Statistical physics suggests that there exists a scaling exponent such that the cost increases as

$$\delta^{\alpha}.$$

New classification of algorithms, intuition suggests that

- low value of $\alpha \Rightarrow$ near-optimal solutions obtained via 'local changes' $\Rightarrow$ algorithmically easy.
- high value of $\alpha \Rightarrow$ near-optimal solutions obtained via long range chages (?) $\Rightarrow$ algorithmically hard.

A TRIVIAL EXAMPLE

Let $(\xi_i, i \geq 1)$ be i.i.d. copies of a strictly positive random variable ξ, and

$$M_n = \max_{A \subseteq \{1,2,\dots,n\}} \sum_{i \in A} (\xi_i - 1) = \max_{A \subseteq \{1,2,\dots,n\}} f(A).$$

with $f(A) = \sum_i 1(i \in A)(\xi_i - 1)$.

The maximum is attained by choosing

$$A_n^{\text{opt}} = \{i \in [n] : \xi_i > 1\}.$$

PERTURBATION OF THE OPTIMAL SET

$$
\begin{array}{lccccccccccccccccccccc}
A_n^{\text{opt}} & \bullet & - & \bullet & - & \circ & - & \bullet & - & \circ & - & \bullet & - & \bullet & - & \circ & - & \bullet & - & \circ & - & \bullet \\
B & \bullet & - & \bullet & - & \circ & - & \bullet & - & \bullet & - & \circ & - & \circ & - & \bullet & - & \bullet & - & \circ & - & \bullet
\end{array}
$$

$\Longrightarrow A_n^{\text{opt}} \triangle B = 4$. Note that a.s.

$$f(A_n^{\text{opt}}) - f(B) > 0.$$

SCALING EXPONENT

We define the random variable:

$$\varepsilon_n(\delta) = \min_{B \subseteq [n]} \left\{ \frac{f(B) - f(A_n^{\text{opt}})}{n} : |B \Delta A_n^{\text{opt}}| \geq \delta n \right\}$$

Do we have, a.s.

$$\varepsilon(\delta) := \lim_n \varepsilon_n(\delta) \sim C\delta^\alpha ?$$

What is the value of α ?

Universality paradigm: α should not depend on the details of the model.

A TRIVIAL EXAMPLE

Now, fix $0 < \delta < 1$ and

$$M_n' = \max_{A' \subseteq [n]} \sum_{i \in A'} (\xi_i - 1)$$

$$\text{subject to } |A' \triangle A_n^{\text{opt}}| \geq \delta n$$

The maximum is attained by $B_n^{\text{opt}} = A_n^{\text{opt}} \triangle D$ where D is the set of indices of the $\lceil \delta n \rceil$ smallest values of $|\xi_i - 1|$.

$$\varepsilon_n(\delta) = n^{-1}(M_n - M_n').$$

A TRIVIAL EXAMPLE

Assume that ξ has a continuous density h. So as $n \to \infty$

$$\varepsilon_n(\delta) \to_{L_1} \varepsilon(\delta) := \int_{1-a(\delta)}^{1+a(\delta)} |x-1| h(x)\, dx$$

where $a(\delta)$ is defined by

$$\delta = \int_{1-a(\delta)}^{1+a(\delta)} h(x)\, dx.$$

So by continuity of $h(x)$, and assuming $0 < h(1) < \infty$, as $\delta \downarrow 0$ we have

$$a(\delta) \sim \frac{\delta}{2h(1)}; \qquad \varepsilon(\delta) \sim a^2(\delta) h(1) \sim \frac{\delta^2}{4h(1)}$$

$\Rightarrow$ This is the desired "scaling exponent $= 2$" result !

NEAR-MINIMAL SOLUTION IN COMBINATORIAL OPTIMIZATION

Under a suitable probabilistic model, heuristic / simulation suggest that

(i) For the NK-model, near minimal spanning tree, the scaling exponent is 2.

(ii) For the travelling salesman problem, minimal pair matching, the scaling exponent is 3.

$\Longrightarrow$ scaling exponent 2 : algorithmically easy problem

$\Longrightarrow$ scaling exponent 3 : algorithmically hard problem

For (i) we have rigorous proofs. For (ii) open question.

ILLUSTRATIVE EXAMPLE

Let $(\xi_i, i \geq 1)$ be i.i.d. copies of a strictly positive random variable ξ, and

$$M_n = \max_{A \subseteq [n]} \left(|A| - \sum_{i=1}^{n-1} \xi_i 1(i \in A, i+1 \in A) \right) = \max_{A \subseteq [n]} f(A).$$

$\Longrightarrow M_n/n \in [1/2, 1]$.

We may prove that if $\xi_i + \xi_{i+1} < 1$ *for some* $1 \leq i \leq n-2$ *then a.s.* A_n^{opt} *is unique.*

We will assume for simplicity that ξ is an exp(λ) variable.

$\Longrightarrow$ A.s., for n large enough, we may define A_n^{opt} as the unique optimal set.

SCALING EXPONENT

Now, fix $0 < \delta < 1$ and

$$M_n' = \max_{A' \subseteq [n]} f(A')$$

subject to $|A' \triangle A_n^{\text{opt}}| \geq \delta n$

and then set

$$\varepsilon_n(\delta) = n^{-1}(M_n - M_n').$$

Do we have:

$$\lim_n \varepsilon_n(\delta) = \varepsilon(\delta) \sim C\delta^2 \quad ?$$

ANALYSIS OF THE OPTIMUM

Recall $M_n = \max_{A\subseteq\{1,2,\ldots,n\}} \left(|A| - \sum_{i=1}^{n-1} \xi_i 1(i \in A, i+1 \in A)\right)$.

We assume for simplicity that ξ is an exponential variable with intensity λ.

Theorem 1. *Almost surely and in* L^1,

$$\lim_{n\to\infty} \frac{M_n}{n} = c(\lambda) = (1 - e^{-\lambda})^{-1} - \lambda^{-1}.$$

$\longrightarrow \lim_{\lambda\to\infty} c(\lambda) = 1, \lim_{\lambda\to 0} c(\lambda) = 1/2.$

$\Longrightarrow$ proof based on a simple instance of Aldous' formulation of the cavity method.

IDEA OF PROOF: LEFT SIDED RECURSION

Define $V_1^L = 1$, $W_1^L = 0$, and

$$\begin{aligned} V_i^L &= \max_{i \in A \subseteq \{1,\ldots,i-1,i\}} \left(|A| - \sum_{j=1}^{i-1} \xi_j 1(j \in A, j+1 \in A) \right) \\ W_i^L &= \max_{i \notin A \subseteq \{1,\ldots,i-1,i\}} \left(|A| - \sum_{j=1}^{i-1} \xi_j 1(j \in A, j+1 \in A) \right) \\ X_i^L &= V_i^L - W_i^L. \end{aligned}$$

Then $M_n = \max(V_n^L, W_n^L)$, and by induction $0 \leq X_i^L \leq 1$,

$$X_{i+1}^L = 1 - \min(X_i^L, \xi_i).$$

IDEA OF PROOF: RIGHT SIDED RECURSION

Define similarly, $V_{n,n}^R = 1$, $W_{n,n}^R = 0$ and

$$\begin{aligned}
V_{n,i}^R &= \max_{i \in A \subseteq \{i,i+1,\ldots,n\}} \left(|A| - \sum_{j=i}^{n-1} \xi_j 1(j \in A, j+1 \in A) \right) \\
W_{n,i}^R &= \max_{i \notin A \subseteq \{i,i+1,\ldots,n\}} \left(|A| - \sum_{j=i}^{n-1} \xi_j 1(j \in A, j+1 \in A) \right), \\
X_{n,i}^R &= V_{n,i}^R - W_{n,i}^R.
\end{aligned}$$

Recursion as i decreases

$$X_{n,i-1}^R = 1 - \min(X_{n,i}^R, \xi_{i-1}).$$

IDEA OF PROOF: CONSTRUCTION OF A_n^{opt}

$-i-(i+1)-$	absolute benefit	relative benefit	when used
$-\bullet--\bullet-$	$V^L+V^R-\xi$	$X^L+X^R-\xi$	if $\xi<\min(X^L,X^R)$
$-\bullet--\circ-$	V^L+W^R	X^L	if $X^R<\min(X^L,\xi)$
$-\circ--\bullet-$	W^L+V^R	X^R	if $X^L<\min(X^R,\xi)$
$-\circ--\circ-$	W^L+W^R	0	never.

IDEA OF PROOF: STATIONARY INFINITE PROCESS

The recursion rule satisfied by X_i^L and $X_{n,i}^R$ does not depend on n. We may define a stationary process $((X_i^L, \xi_i, X_{i+1}^R), -\infty < i < \infty)$ satisfying the same recursion rule:

$$X_i^L = 1 - \min(X_{i-1}^L, \xi_{i-1})$$
$$X_i^R = 1 - \min(X_{i+1}^R, \xi_i).$$

This type of recursion equation is called a Recursive Distribution Equation (RDE).

$\Longrightarrow$ For each $i \in \mathbb{Z}$, X_i^L, ξ_i and X_{i+1}^R are independent variables.

$\Longrightarrow$ It is possible to compute the stationary distribution of this process.

$\Longrightarrow$ We define an infinite set A^{opt} using the construction table.

IDEA OF PROOF: LOCAL WEAK CONVERGENCE

Lemma 1. *Let* U_n *be uniform on* $\{1,\ldots,n\}$*. As* $n \to \infty$

$$((X^L_{U_n+i}, \xi_{U_n+i}, X^R_{n,U_n+i+1}, 1(U_n + i \in A_n^{\text{opt}})),\ -\infty < i < \infty)$$

$$\xrightarrow{d}\ ((X^L_i, \xi_i, X^R_{i+1}, 1(i \in A^{\text{opt}})),\ -\infty < i < \infty)$$

where the left side is defined arbitrarily for $U_n + i \notin \{1,\ldots,n\}$ *and where convergence in distribution is with respect to the usual product topology on infinite sequence space.*

IDEA OF PROOF: IDENTIFICATION OF THE LIMIT

Recall that $M_n = \max_{A\subseteq\{1,2,\ldots,n\}} \left(|A| - \sum_{i=1}^{n-1} \xi_i 1(i \in A, i+1 \in A)\right)$.

The local weak convergence lemma implies that

$$c(\lambda) = \lim_{n\to\infty} \mathbb{E}\frac{M_n}{n} = \mathbb{P}(0 \in A^{\text{opt}}) - \mathbb{E}\xi 1(0 \in A^{\text{opt}}, 1 \in A^{\text{opt}}).$$

NEAR MINIMAL SOLUTION

Recall that $f(A) = |A| - \sum_i \xi_i 1(i \in A, i+1 \in A)$ and

$$\varepsilon_n(\delta) = \min_{B \subseteq \{1,\cdots,n\}} \left\{ \frac{f(B) - f(A_n^{\text{opt}})}{n} : |B \Delta A_n^{\text{opt}}| \geq \delta n \right\}.$$

Theorem 2. $\varepsilon(\delta) = \lim_n \mathbb{E}\varepsilon_n(\delta)$ *exists for all* $0 < \delta < 1$, *and*

$$\limsup_{\delta \downarrow 0} \delta^{-2} \varepsilon(\delta) < \infty,$$

and

$$\liminf_{\delta \downarrow 0} \delta^{-2} \varepsilon(\delta) > 0.$$

$\Longrightarrow$ Scaling exponent is 2.

IDEA OF PROOF: LAGRANGE MULTIPLIER

We introduce the Lagrange multiplier $\theta > 0$, and

$$\max_{B \subseteq \{1,\dots,n\}} \left(|B| - \sum_{i=-\infty}^{\infty} \xi_i 1(i \in B, i+1 \in B) + \theta |B \triangle A_n^{\text{opt}}| \right),$$

and define $B_n^{\text{opt}}(\theta)$ as the corresponding optimizing set.

We extend our previous analysis to this new optimization problem.

We define a new pair of random variables $(Z^L(\theta), Z^R(\theta))$ playing the role of (X^L, X^R).

$\longrightarrow X^{L,R} = Z^{L,R}(0)$ and $A_n^{\text{opt}} = B_n^{\text{opt}}(0)$.

IDEA OF PROOF: QUINTUPLE PROCESS

We define a stationary quintuple process
$((Z_i^L(\theta), X_i^L, \xi_i, X_{i+1}^R, Z_{i+1}^R(\theta)), -\infty < i < \infty)$.

$\longrightarrow$ The infinite sets A^{opt} and $B^{\text{opt}}(\theta)$ are built thanks to the quintuple process according to a construction table, and

$$\{i \in B^{\text{opt}}(\theta)\} \text{ is } \sigma((Z_i^L(\theta), X_i^L, \xi_i, X_{i+1}^R, Z_{i+1}^R(\theta))\text{-measurable.}$$

Let $J_i = 1(i \notin A^{\text{opt}}) - 1(i \in A^{\text{opt}})$, we find

$$\begin{aligned} Z_i^L &= 1 - \min(Z_{i-1}^L, \xi_{i-1}) 1(Z_{i-1}^L \geq 0) + \theta J_i \\ Z_i^R &= 1 - \min(Z_{i+1}^L, \xi_i) 1(Z_{i+1}^L \geq 0) + \theta J_i \end{aligned}$$

$\longrightarrow$ A new RDE for (Z^L, Z^R) that we cannot solve analytically.

IDEA OF PROOF: NEAR-MINIMAL SOLUTIONS

The proportion of items at which A^{opt} and B^{opt} differ is

$$\begin{aligned}\delta(\theta) &= \mathbb{P}(\{0 \in A^{\text{opt}}\}\triangle\{0 \in B^{\text{opt}}\}) \\ &= \mathbb{P}\left(\{X^L > \min(X^R, \xi)\}\triangle\{Z^L - \theta J_0 > \min((Z^R - \theta J_1)^+, \xi)\}\right)\end{aligned}$$

and the difference in mean benefit per item between A^{opt} and B^{opt} is

$$\begin{aligned}\bar{\epsilon}(\theta) &= \mathbb{E}[1(0 \in A^{\text{opt}}) - \xi 1(0 \in A^{\text{opt}}, 1 \in A^{\text{opt}})] \\ &\qquad - \mathbb{E}[1(0 \in B^{\text{opt}}) - \mathbb{E}\xi 1(0 \in B^{\text{opt}}, 1 \in B^{\text{opt}})] \\ &= \mathbb{E}[\texttt{complicated expression}].\end{aligned}$$

We have

$$\lim_{n\to\infty} \mathbb{E}\varepsilon_n(\delta(\theta)) = \bar{\epsilon}(\theta).$$

IDEA OF PROOF: VARIATIONAL ANALYSIS

We write

$$Z^{L,R} - \theta J_0 = X^{L,R} + \theta D^{L,R}(\theta).$$

Define $X = (X^L, \xi, X^R)$ and $D(\theta) = (D^L(\theta), 0, D^R(\theta))$, then

$$\begin{aligned}\delta(\theta) \;&=\; \mathbb{P}(X \in \Sigma, X + \theta D(\theta) \notin \Sigma) \\ &\quad + \mathbb{P}(X + \theta D(\theta) \in \Sigma, X \notin \Sigma)\end{aligned}$$

with

$$\Sigma = \left\{(x^L, t, x^R) \in \mathbb{R}^3 : x^L > \min(t, \max(0, x^R))\right\}.$$

IDEA OF PROOF: VARIATIONAL ANALYSIS

$\Rightarrow$ *Let* $F(\theta) = \mathbb{P}(X \in \Sigma, X + \theta D(\theta) \notin \Sigma)$*, we prove that*

$$\delta(\theta) \sim 2F'(0)\theta$$

and there is an integral expression for $F'(0)$*.*

$\Longrightarrow$ Similarly, we could try to check that $\bar{\epsilon}(\theta) \sim G''(0)\theta^2$.

However computation is hard...

We using a totally different probabilistic argument, we prove a weaker

$$C_1\theta^2 \le \bar{\epsilon}(\theta) \le C_2\theta^2.$$

IDEA OF PROOF: UPPER AND LOWER BOUND

- *UPPER BOUND:*

  $\limsup \delta^{-2}\varepsilon(\delta) < \infty$*:* we identify a local configuration ● ○ ● ○ ● ○ ● on $\{i, \cdots, i+K\}$which can be replaced by ● ● ○ ● ○ ● ● at a low extra cost $1 - \xi_i - \xi_{i+K-1}$.

- *LOWER BOUND:*

  $\liminf \delta^{-2}\varepsilon(\delta) > 0$*:* we prove that $C_1\theta^2 \leq \bar{\epsilon}(\theta)$ by expressing $\bar{\epsilon}(\theta)$ as a mean cost increase over a block of finite random length. We then lower bound the mean increase cost if B^{opt} and A^{opt} differ over a block by a combinatorial argument.

SCALING OF THE LAGRANGE PARAMETER

For the *NK-model / minimal spanning tree of a Poisson point process on* $\mathbb{R}^d$:

$$\delta(\theta) \asymp \theta \quad \text{and} \quad \bar{\epsilon}(\theta) \asymp \theta^2.$$

$\longrightarrow$ scaling exponent 2.

Under a suitable probabilistic model, for the *traveling salesman problem / minimal pair matching*, simulation suggests [Aldous and Percus (2003)]

$$\delta(\theta) \asymp \theta^{1/2} \quad \text{and} \quad \bar{\epsilon}(\theta) \asymp \theta^{3/2}.$$

$\longrightarrow$ scaling exponent 3.

There exists a related RDE for these optimization problems [Aldous (2001)] but no mathematical explanation of this phenomenon.

THE KAUFFMAN-LEVIN NK-MODEL

The Kauffman-Levin NK model of random fitness landscape has attracted extensive literature in statistical physics.

Let $K \geq 2$ and

$$M_n = \max_{A \subseteq [n]} \sum_{i \in A} -W_i(A_{[i,i+K]}),$$

where $A_{[i,i+K]}$ denotes the set A restricted on the interval $\{i, \cdots, i+K\}$ and

$$(W_i(B), i \geq 1, B \subseteq [K+1])$$

are independent exp(1) random variables.

$\Rightarrow$ This is algorithmically easy via dynamic programming.

$\Rightarrow$ By Kingman's Subadditive Theorem there is an a.s. limit $n^{-1} M_n \to c_K$.

THE KAUFFMAN-LEVIN NK-MODEL: NEAR MINIMAL SOLUTION

Now, fix $0 < \delta < 1$ and

$$M'_n = \max_{A' \subseteq [n]} \sum_{i \in A'} -W_i(A_{[i,i+K]})$$

$$\text{subject to } |A' \triangle A_n^{\text{opt}}| \geq \delta n$$

and then set

$$\varepsilon_n(\delta) = n^{-1}(M_n - M'_n)$$

We expect the existence of the a.s. limit

$$\varepsilon(\delta) = \lim_{n \to \infty} \varepsilon_n(\delta)$$

and simulation and heuristics suggest that, as $\delta \downarrow 0$

$$\varepsilon(\delta) \sim C\delta^2.$$

$\Rightarrow$ Again "Scaling exponent $= 2$" !