Correlated Geometric Mutations for Integer Evolution Strategies

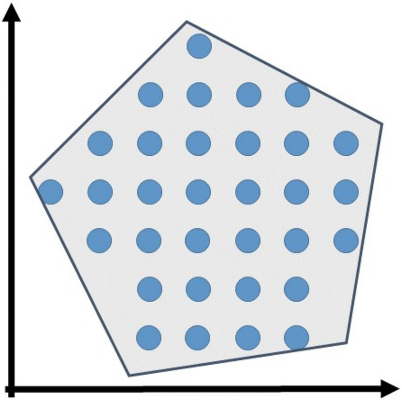
Ofer M. **Shir** (Tel-Hai College & Migal Institute, IL)
Michael **Emmerich** (University of Jyväskylä, FI)

The Genetic and Evolutionary Computation Conference — GECCO'25, AABOH workshop, July 2025, Málaga, Spain

https://doi.org/10.1145/3712255.3734283

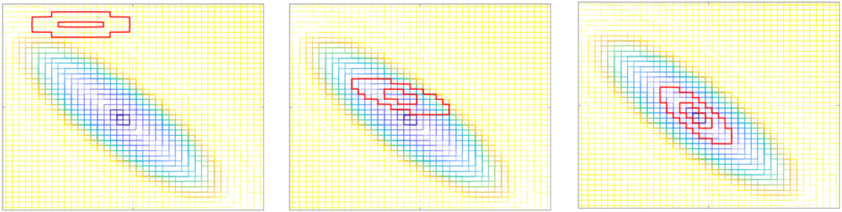
Domain: Integer Evolution Strategies (IESs)

We are interested in IESs for their (i) intrinsic **mixed-integer capabilities**, (ii) well-developed **self-adaptation mechanisms**, and (iii) high efficacy in handling **unbounded search spaces**.



status & questions

Existing IESs work well, usually by applying the **Truncated Normal** (TN) distribution in their mutation operator:



- But *no questions asked* on the mutations' behavior.
- Rudolph [1994] identified the Double-Geometric (DG) distribution as a promising tool for uncorrelated integer mutations.
- **Questions**: (i) Are we able to well-define correlated DG-driven mutations, and if so, (ii) will they be beneficial?

preliminaries

TN:

univariate – $z_0 \sim \mathcal{N}\left(0, \sigma^2\right) \quad \Longrightarrow \quad z = \texttt{round}\left(z_0\right)$

multivariate – $\vec{z}_0 \sim \mathcal{N}\left(\vec{0}, \mathbf{C}\right) \quad \Longrightarrow \quad \vec{z} = \texttt{round}\left(\vec{z}_0\right)$

DG:

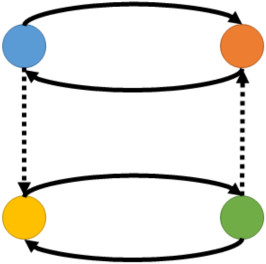
univariate – $g_i \sim \mathcal{G}\left(0, p\right) \;\; (i = 1, 2) \quad \Longrightarrow \quad z = \left(g_1 - g_2\right)$

multivariate – i.i.d. of the above: $z_j = \mathcal{G}\left(0, p_j\right) - \mathcal{G}\left(0, p_j\right) \quad j = 1 \ldots n$

correlated multivariate – unknown

The DG distribution is controlled by the ℓ_1-norm-driven *mean step-size*, $S = \mathbb{E}\left[\|\vec{z}\|_1\right] = \sum_{i=1}^{n_z} \mathbb{E}\left[|z_i|_1\right]$ (due to the stochastic independence):

$$p = 1 - \frac{S/n_z}{\sqrt{\left(1 + (S/n_z)^2\right)} + 1} \quad \Longleftrightarrow \quad S = n_z \cdot \frac{2(1-p)}{p(2-p)}.$$



methodology

```
/* This function generates uncorrelated random vectors
according to either the Geometric or Normal distributions. */
ies::genUncorrelatedMutation(σ⃗, type)
    n ⟵ len(σ⃗),    z⃗ := 0⃗ ∈ ℝ^n
    if type==DG then
        /* default Geometric */
        for i = 1, ..., n do
            p_i ⟵ 1 − (σ_i/n) / (sqrt(1 + (σ_i/n)^2) + 1)
            z_i ⟵ 𝒢(0, p_i)
        end
    else
        /* default Normal */
        for i = 1, ..., n do
            z_i ⟵ σ_i · 𝒩(0, 1)
        end
    end
    return {z⃗}
```

$$n \longleftarrow \texttt{len}(\vec{\sigma}), \quad \vec{z} := \vec{0} \in \mathbb{R}^n$$

$$p_i \longleftarrow 1 - \frac{\sigma_i/n}{\sqrt{\left(1+(\sigma_i/n)^2\right)}+1}$$

$$z_i \longleftarrow \mathcal{G}(0, p_i)$$

$$z_i \longleftarrow \sigma_i \cdot \mathcal{N}(0,1)$$

Schwefel's rotations (i)

We capitalize on Schwefel's definition of the standard ES, according to which the covariance information is stored by means of the n-dimensional variances' vector $\vec{\sigma}$ as well as the $n(n-1)/2$-dimensional vector of rotational angles $\vec{\alpha}$.

The transformation of a covariance element c_{ij} into a rotational angle α_{ij} (where $c_{ii} \equiv \sigma_i^2$) provides a useful relationship for decision variables i and j:

$$\alpha_{ij} = \frac{1}{2} \arctan\left(\frac{2c_{ij}}{\sigma_i^2 - \sigma_j^2}\right) ,$$

where $\alpha_{ij} = 0$ whenever no correlation exists.

Schwefel's rotations (ii)

The realization of the correlated mutation instance $\vec{z}_c$ is achieved by a sequence of $n(n-1)/2$ rotations using the operator $\mathbf{R}(\theta) := (r_{k\ell})$

$$\vec{z}_c = \left(\prod_{i=1}^{n-1} \prod_{j=i+1}^{n} \mathbf{R}(\alpha_{ij}) \right) \cdot \vec{z}_u \ . \tag{1}$$

$\mathbf{R}$'s matrix form is identical to the unity, except for 4 elements:

$$r_{kk} = r_{\ell\ell} = \cos(\alpha_{k\ell}), \qquad r_{k\ell} = -r_{\ell k} = -\sin(\alpha_{k\ell}).$$

Rudolph [1992] verified the validity of this representation.

$\texttt{rotate}\,(\vec{z},\ \vec{\alpha})$
  for $j = 1, \ldots, n \cdot (n-1)/2$ **do**
    $\vec{z} \longleftarrow \mathbf{R}(\alpha_j)\vec{z}$
  end
  return $\{\vec{z}\}$

```
ies::corrMutate(x⃗, σ⃗, α⃗, n, type)
```

$\mathcal{N}_g \leftarrow \mathcal{N}(0,1)$, $\tau_g \leftarrow \frac{1}{\sqrt{2\cdot n}}$, $\tau_\ell \leftarrow \frac{1}{\sqrt{2\cdot\sqrt{n}}}$

for $i = 1, \ldots, n$ **do**

  $\sigma'_i \longleftarrow \sigma_i \cdot \exp\{\tau_g \cdot \mathcal{N}_g + \tau_\ell \cdot \mathcal{N}_i(0,1)\}$

end

for $j = 1, \ldots, n \cdot (n-1)/2$ **do**

  $\alpha'_j \longleftarrow \alpha_j + \beta \cdot \mathcal{N}_j(0,1)$

end

$\vec{z}_u \longleftarrow$ `genUncorrelatedMutation` $(\vec{\sigma}', \texttt{type})$

$\vec{z} \longleftarrow$ `round` $($`rotate` $(\vec{z}_u, \vec{\alpha}'))$ /* ROTATE & ROUND ! */

if *type==DG* **then**

  $\vec{z}_g \longleftarrow$ `genUncorrelatedMutation` $(\vec{\sigma}', \texttt{type})$

  $\vec{z}'_g \longleftarrow$ `round` $($`rotate` $(\vec{z}_g, \vec{\alpha}'))$

  $\vec{z} \longleftarrow \vec{z} - \vec{z}'_g$ /* difference of two geometric samples */

end

$\vec{x}' \longleftarrow \vec{x} + \vec{z}$

return $\{\vec{x}', \vec{\sigma}', \vec{\alpha}'\}$

(μ, λ) Integer Evolution Strategy

$t \leftarrow 0$
$P(t) \leftarrow$ `randIntUniform`(μ) /* forming μ individuals, each with decision variables $\vec{x}$ + strategy parameters $\{\vec{\sigma}, \vec{\alpha}\}$ */
`evaluate`$(P(t))$
repeat
  $P'(t) \leftarrow$ `recombine`$(P(t))$ /* forming λ offspring by repeatedly drawing $\frac{\lambda}{2}$ pairs of parents at random */
  $P''(t) \leftarrow$ `mutate`$(P'(t),$ `type`$)$ /* calling `corrMutate`, which also self-adapts the strategy parameters */
  `evaluate`$(P''(t))$
  $P(t+1) \leftarrow$ `select`$(P''(t))$ /* deterministically selecting the top μ individuals post-sorting */
  $t \leftarrow t + 1$
until *evaluation budget is exhausted*
return { *best individual found* }

2D populations

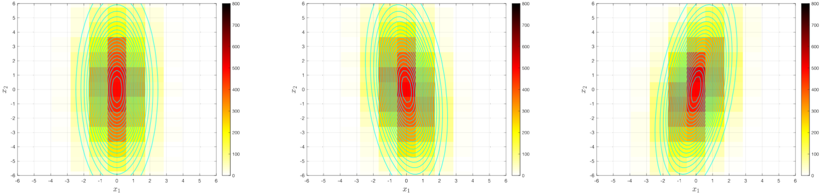
We present heatmaps of both TN- and DG-based 2D sampled populations of size 10^4 per $\sigma_1 = 1.0$ and $\sigma_2 = 3.0$: uncorrelated (diagonal), correlated (nondiagonal) with $c_{12} = -0.8$, and with $c_{12} = 1.2$ (assuming a structure of the form $[\sigma_1^2, c_{12}; c_{12}, \sigma_2^2]$).

Since the simulation is governed by the Normal distribution's parameters, the DG's step-size can be approximated as

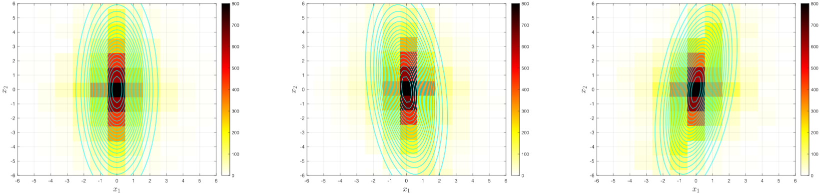
$$S_i \approx \int_{-\infty}^{\infty} |z| \cdot \mathrm{pdf}(z)\ \mathrm{d}z = \sigma_i \cdot \sqrt{\frac{2}{\pi}}$$

2D visualzation

TN:



DG:



`numerical observations`

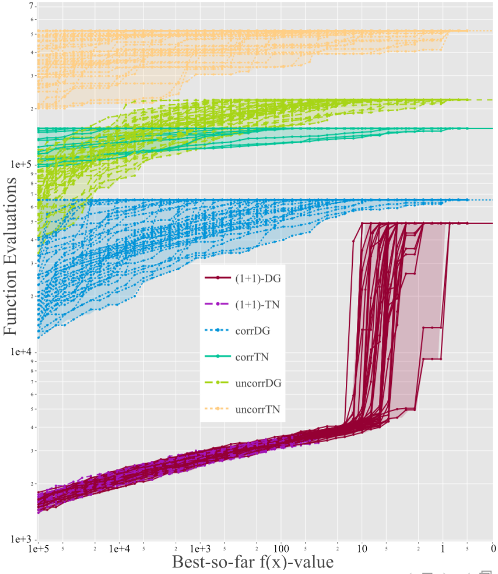
preliminary: (1+1)-IES on the Integer Sphere

$$\begin{array}{ll} \text{minimize}_{\vec{x}} & \vec{x}^T \vec{x} \\ \text{subject to:} & \vec{x} \in \mathbb{Z}^n \end{array}$$

We utilize Rechenberg's renowned **1/5th success-rule** for the step-size adaptation, in play with either the TN or DG mutation distributions, and compare six strategies:

1. (1+1)-DG
2. uncorrelatedDG
3. correlatedDG
4. (1+1)-TN
5. uncorrelatedTN
6. correlatedTN

six IESs over the 80D Integer Sphere



unbounded integer quadratic optimization problems

We seek numerical validation to our hypotheses by considering unbounded quadratic integer optimization problems of the following class:

$$\boxed{\begin{array}{ll} \text{minimize}_{\vec{x}} & \frac{1}{c} \cdot \left[\left(\vec{x} - \vec{\xi}_0 \right)^T \mathbf{H} \left(\vec{x} - \vec{\xi}_0 \right) \right] \\ \text{subject to:} & \vec{x} \in \mathbb{Z}^n, \end{array}}$$

where the Hessian matrix $\mathbf{H}$, its parametric condition number c and the location vector $\vec{\xi}_0$ completely define a problem instance.

IQP instances

We consider 4 $n \times n$ Hessian matrices to represent two separable (i.e., diagonal forms) and two nonseparable (i.e., nondiagonal forms) problems:

H-1 Discus: $(\mathcal{H}_{\text{disc}})_{11} = c, \;\; (\mathcal{H}_{\text{disc}})_{ii} = 1 \quad i = 2, \ldots, n;$

H-2 Cigar: $(\mathcal{H}_{\text{cigar}})_{11} = 1, \;\; (\mathcal{H}_{\text{cigar}})_{ii} = c \quad i = 2, \ldots, n;$

H-3 Rotated Ellipse (RotEllipse):

$$\mathcal{H}_{\text{RE}} = \mathcal{O}\mathcal{H}_{\text{ellipse}}\mathcal{O}^{-1}$$

where $\mathcal{O}$ is rotation by $\approx \frac{\pi}{4}$ radians in the plane spanned by $(1, 0, 1, 0, \ldots)^T$ and $(0, 1, 0, 1, \ldots)^T$;

H-4 Hadamard Ellipse (HadEllipse):

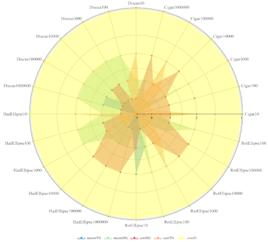
$$\mathcal{H}_{\text{HE}} = \mathcal{S}\mathcal{H}_{\text{ellipse}}\mathcal{S}^{-1}$$

where the rotation constitutes the normalized Hadamard matrix, $\mathcal{S} := \text{Hadamard}(n)/\sqrt{n}$.

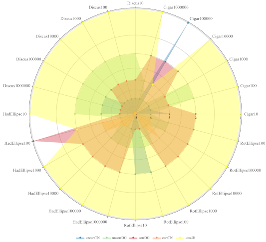
We consider 6 levels of conditioning, $c \in \{10, 10^2, \ldots, 10^6\}$, which yield altogether 24 problem instances per dimensionality.

overall performance when considering also the `cmaIH`

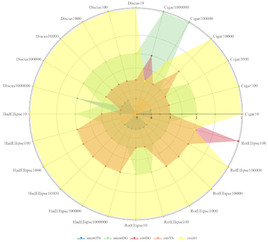
32D



64D

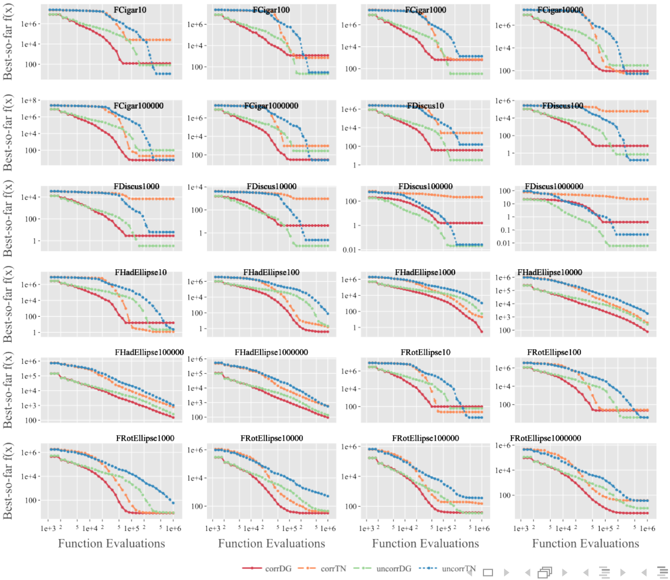


128D



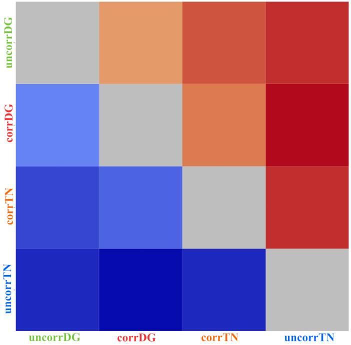
The ranking of the five IESs (including the `cmaIH`) using radar charts across the 24 problem instances (serving as nodes). The performance is ranked using fixed-budget analyses (with “rank-1” being the winner).

fixed-budget gallery per 64D

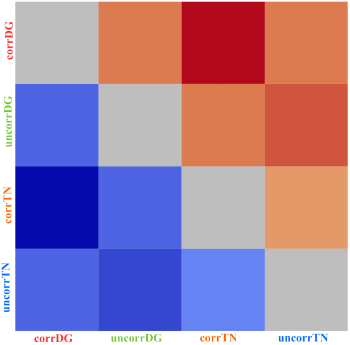


pairwise numerical comparisons amongst the four IESs

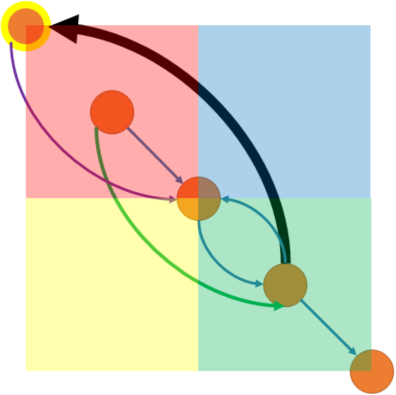
64D: Separable



64D: Nonseparable



`uncorrDG` dominates the **separable** subset (`corrDG` is second); `corrDG` dominates the **nonseparable** subset (`uncorrDG` is second) – consistently across dimensions (see 32D and 128D in the paper).



discussion

summary

- We proposed a procedure for generating correlated DG mutations.
- We showed that the $(1+1)$-IES with DG mutations worked well with the $1/5$th success-rule on the unconstrained integer Sphere model without any adjustments, unlike its TN-based counterpart.
- Concerning the IQP test-suite:
  - DG-based IESs always outperform TN-based IESs over the tested suite.
  - Correlated DG mutations are beneficial per the tested nonseparable IQP problems.
- See also a recept paper by Zepko & Shir, "All-Quadratic Mixed-Integer Problems: A Study on Evolution Strategies and Mathematical Programming" (accepted to ECJ).

take-home messages

- The DG distribution should be further investigated:
  - to the adaptation framework of the derandomized CMA-ES;
  - extended analysis over a wider range of model-landscapes;
  - statistical properties of the correlated DG, e.g., entropy, might reveal important insights
- What mechanism enables the `cmaIH` to outperform the other IESs? We now understand that Gaussianity does not give an advantage. We *speculate* that the advanced step-size control mechanism is responsible for that.
- Coming-up at FOGA'25: a fundamental study with a rigorous investigation of the two mutation distributions – Shir & Emmerich, "Foundations of Correlated Mutations for Integer Programming".

Ofer Shir acknowledges the Faculty of Biology at the Technion - Israel Institute of Technology for hosting him during his sabbatical.

gracias