Foundations of Correlated Mutations for Integer Programming

Ofer M. Shir
Tel-Hai College and Migal Institute
Upper Galilee, Israel
ofersh@telhai.ac.il

Michael Emmerich
University of Jyväskylä
Finland
michael.t.m.emmerich@jyu.fi

ABSTRACT

Even with the recent theoretical advancements that dramatically reduced the complexity of Integer Programming (IP), heuristics remain the dominant problem-solvers for this difficult category. This study seeks to establish the foundation for Evolution Strategies (ESs), a class of randomized search heuristics inherently designed for continuous spaces. We particularly focus on ESs for their intrinsic mixed-integer capabilities, well-developed self-adaptation mechanisms, and high efficacy in handling unbounded search spaces. ESs already excel in treating IP in practice, but accomplish it via discretization and by applying sophisticated patches to their continuous operators, while persistently using the ℓ_2-norm as their operation pillar. We lay foundations for discrete search by adopting the ℓ_1-norm, accounting for the suitable step-size, and exploring mutation distributions for unbounded integer decision variables. We narrow down to the Truncated Normal (TN) and Double Geometric (DG) distributions, explore their theoretical properties, including entropy functions, and propose a procedure to generate scalable correlated mutations. Our investigations are accompanied by extensive numerical simulations, which consistently support the claim that the DG distribution is better suited for unbounded integer search. We link our theoretical perspective to empirical evidence indicating that an ES with correlated DG mutations outperforms other strategies over non-separable quadratic IP. We conclude that substituting the default TN distribution with the DG is theoretically justified and practically beneficial. We also advocate for the use of the ℓ_1-norm over the ℓ_2-norm based on the empirical indications presented, although it remains to be formally proven and/or rigorously benchmarked.

KEYWORDS

Integer Programming, multivariate discrete distributions, Truncated Normal, Double Geometric, evolution strategies, ℓ_1-norm

1 INTRODUCTION

Integer Programming (IP) comprises a rich family of optimization problems with discrete decision variables and is central to both practical computation and the theoretical foundations of Computer Science [17, 23]. Whereas IP is often viewed as a generalization of linear programming for combinatorial optimization, our paper focuses on methods for unconstrained or loosely constrained problems [26]. As opposed to categorical integer variables, we consider metric (or *ordinal*) integer variables – whose objective-function values are correlated with the ℓ_1 (generalized Hamming) distance among decision variable vectors. Applications include selecting pipeline diameters from a discrete set of options [19] and determining the number of stages in separation processes within chemical plants [6]. Metric integer parameters also arise in hyperparameter tuning for machine learning, image processing, and inventory planning, where decisions must be expressed as integer values [15].

However, tackling such integer optimization problems remains computationally formidable in the worst case. The *decision* version of IP is $\mathcal{NP}$-hard, and for decades the fastest exact algorithm was due to Kannan, with running time $\text{poly}(n)\, 2^{O(n)}$ for an instance with n variables [13]. A recent breakthrough by Rothvoss and Reis tightened this bound: they presented a randomized algorithm that solves any IP in $(\log(2n))^{O(n)}$ steps [20]. Despite this major theoretical advancement and the formulation of a faster algorithm, heuristics remain the primary tool to address IP problem-solving in practice. In particular, there is a need for solvers that can work in an opaque-box setting, that is, without explicitly knowing the mathematical expression of the objective functions. Also in other aspects, unconstrained IP significantly differs from its continuous counterpart. For example, a convex quadratic optimization problem can become multimodal (i.e., with local traps) when discretized – as we demonstrate in the Supplementary Material.

When it comes to heuristic methods for solving unconstrained optimization problems, Evolution Strategies (ESs) [2, 18, 24] are effective randomized search heuristics. They stand out from many other Evolutionary Algorithms due to search mechanisms motivated by a blend of inspiration from biological evolution and a thorough theoretical analysis [4], advanced step-size adaptation mechanisms [1], and their invariance to monotonic transformations of the objective function as well as rotations in the decision space [9]. ESs are commonly used in continuous spaces, but their paradigms have been adapted to treat integer [22] and mixed-integer (MI) spaces (see, e.g., [3, 6, 15, 26]). So far, these adaptations primarily rely on truncated versions of mutation distributions originally designed for continuous settings. For instance, modern ESs consistently demonstrate strong capabilities in addressing MI optimization problems, while still operating with the Truncated Normal (TN) distribution under-the-hood after certain adaptations (e.g., the CMA-ES-wM [8] or the CMA-ES-IH [16]). Despite this ability of modern ESs to effectively solve such problems, key theoretical questions remain unanswered. This practical success has, so far, offered little incentive to explore these open issues – particularly the suitability of alternative discrete distributions for integer mutations, where the effectiveness of the TN distribution remains unclear. Meanwhile, other areas of MI foundations in ESs have recently attracted growing interest (see, e.g., [12, 30]).

FOGA '25, August 27–29, 2025, Leiden, Netherlands

ACM ISBN 979-8-4007-1859-5/2025/08
https://doi.org/10.1145/3729878.3746698

Rudolph, following the principle of maximum entropy, justified the use of the Double Geometric (DG) distribution for unbounded integer optimization in a singular study published in 1994 [22]. Rudolph proposed therein a mutation operator for integer search-spaces, grounded in statistical theory, using an independent multivariate sample from the DG distribution. Following it, the DG distribution has been utilized in MI ESs (MIESs) [6, 15], with empirical evidence demonstrating that mutative step-size control can effectively track nearly optimal step-sizes for the so-called Progress Rate [15]. However, such studies with DG-based mutations have remained isolated examples compared to the multivariate TN distribution, which has been performing successfully and, therefore, like the "winning horse in midstream", has kept its steady state as the first choice for the designers and practitioners of MIESs. In this context, a recent empirical study [5] demonstrated that Gaussianity is not essential for ESs' strong performance over *continuous* spaces, motivating further exploration of alternative distributions. In addition to this momentum effect, another reason for persisting with the TN distribution might be that a multivariate correlation structure has never been defined for the DG distribution, and it is an open question to which extent such a correlation structure can be imposed. This key question will drive us to investigate the foundations of heuristic search over the integer lattice, while conducting extensive simulations for validation.[1] Moreover, unlike the continuous domain which enjoys direct ℓ_2-norm-driven relations between the statistical covariance to translation and rotation transformations, the existence of such relations or even proper definitions of non-trivial rotations over the integer lattice remains an open issue. In the context of ESs' research, the required norm-shift from ℓ_2 to ℓ_1 constitutes a major challenge with a limited ability to capitalize on the large volume of well-established ℓ_2-driven results.

Research Questions. Motivated by the differences in geometric structure between continuous and integer spaces and by empirical observations on mutation behavior, this paper investigates the following research questions:

(1) **Integer Mutation Distributions: Geometry & Entropy.**
   - **Open Question:** How can mutation operators for integer ESs (IESs) be designed to respect the inherent ℓ_1-norm (Manhattan) geometry of the integer lattice? How does its entropy and symmetry compare to other proposed distributions, such as the TN distribution?
   - **Hypothesis:** For integer optimization problems ℓ_1-invariant operators are better suited than ℓ_2-invariant ones. Moreover, the mean ℓ_1 metric is a more meaningful measure of the average step-size than the standard deviation.

(2) **Correlated Integer Mutations via DG Distributions.**
   - **Open Question:** Can a correlated version of the DG distribution be constructed that accurately preserves intended dependencies among integer variables? What is the exact formulation and what are the properties?
   - **Hypothesis:** A sampling scheme based on the DG distribution can be extended to the correlated case while preserving covariance structure and yielding higher entropy than truncated Gaussians. The DG distributions, which can be adjusted to a given covariance structure in a rather straightforward manner, is not well suited to reflect ℓ_1 – symmetries, which might make it less suitable for unconstrained integer optimization.

(3) **Stagnation near Optima: Discreteness-Induced Phase Transition.**
   - **Open Question:** Why do integer-based ESs systematically stagnate near the optimum, and can mutation dynamics be adapted to mitigate this behavior?
   - **Hypothesis:** A phase-transition-like stagnation effect occurs due to the reduced move options in the near-optimal integer lattice; standard adaptation strategies fail to overcome it without explicitly accounting for discreteness.

Paper Organization. We provide the mathematical preliminaries in Section 2, where we also state our working hypothesis and present basic numerical assessment of the preliminaries. Section 3 is dedicated to the rotation transformation for generating correlated mutations, while Section 4 covers the *entropy* perspective of the investigated mutation distributions. In Section 5 we present numerical observations of solving integer quadratic programs as a proof-of-concept. Finally, we summarize our work in Section 6.

2 PRELIMINARIES

The purpose of this section is to present the necessary preliminaries, including our working hypotheses, and then describe four discrete distributions of interest, namely the Discrete Uniform, the Shifted Binomial, the Truncated Normal and the Double Geometric.

2.1 Framework and Notation

The general framework that underlies our work is MI optimization problem-solving, subscribing to the following formulation:

$$\begin{array}{ll} \text{minimize}_{\vec{y}} & f(\vec{y}) \\ \text{subject to:} & g(\vec{y}) \\ & \vec{y} \in \mathbb{R}^D \\ & y_i \in \mathbb{Z} \quad \forall i \in I, \end{array} \tag{1}$$

where f and g constitute the objective and constraint functions, respectively. Importantly, the D-dimensional decision vector $\vec{y}$ is constructed by n_r real-valued decision variables followed by n_z (metric) integer decision variables that are defined by the so-called *index set* $I := \{n_r + 1, \ldots, n_r + n_z\}$: $\forall i \in I \quad y_i \in \mathbb{Z}$. As mentioned earlier, our consideration excludes categorical decision variables, although they may as well exist in practical modeling (see, e.g., [27]) – but exceed the scope herein. While Eq. (1) defines the context and induces the applicability and relevance of this study thereafter, **we focus entirely on the integer subspace, and thus refer to the integer decision vector as our target, $\vec{x} \in \mathbb{Z}^n$, fully mapped onto the index set I and satisfying $n = n_z$.** Moreover, given the context of Eq. (1), our focus on ESs is further justified by their MI capabilities, well developed self-adaptation mechanisms, and high efficacy in handling unbounded search spaces. In ESs' operation, during the solution process, an existing integer decision vector $\vec{x}_{\text{CURR}}$ undergoes variations to generate a candidate vector $\vec{x}_{\text{NEW}}$:

$$\vec{x}_{\text{NEW}} = \vec{x}_{\text{CURR}} + \vec{z}.$$

[1] The source code supporting this research is publicly available at https://github.com/ofersh/correlated-mutations-for-integer-programming .

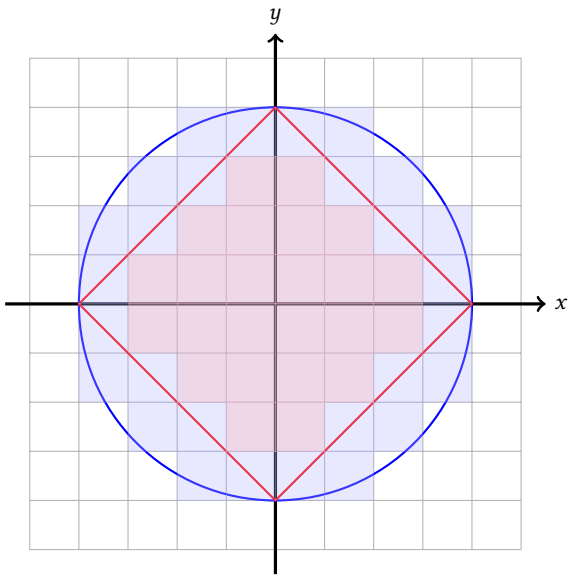

Figure 1: A visual comparison of the ℓ_1-norm versus the ℓ_2-norm over the 2D integer lattice using spheres of integer radius 4: an ℓ_1-sphere (red) versus an ℓ_2-sphere (blue).

$\vec{z} \in \mathbb{Z}^n$ is the mutation vector, which is drawn from a certain multivariate distribution. This distribution may possess a correlated or uncorrelated nature with respect to the generated random variables (that is, with respect to the independence of z_i). Unless specified otherwise, we assume here uncorrelated distributions.

To assess the mutation vectors, we utilize the ℓ_1-norm, which reflects the vector length over the integer lattice since it quantifies the minimal number of elementary steps required to reach one vector from another:

$$\|\vec{z}\|_1 := \sum_{i=1}^{n} |z_i|. \tag{2}$$

See Figure 1 for an illustrated comparison of spheres (or disc boundaries) induced by the two norms in 2D, when visualized over the same integer lattice. For positive integer radii, the set of points at a fixed ℓ_1 distance from the origin forms a *discrete sphere* (or "diamond") that lies entirely within the integer lattice. These ℓ_1-spheres are fully representable, meaning that every point on the sphere has integer coordinates, which is generally not true for ℓ_2-spheres.

The expected ℓ_1-norm constitutes the mean step-size,

$$S := \mathbb{E}\left[\|\vec{z}\|_1\right] = \sum_{i=1}^{n} \mathbb{E}\left[|z_i|_1\right], \tag{3}$$

due to the stochastic independence. When the n random variables are identically drawn under the same settings, this mean step-size becomes $S = n \cdot \mathbb{E}\left[|z_1|\right]$.

The probability distributions under consideration are discrete and they are formulated by means of the probability of the mutation element to take an integer value k:

$$\mathbf{Pr}\{z = k\} = p_k, \quad \sum_k p_k = 1.0.$$

Finally, we mention Shannon's entropy as a measure of averaged information, which quantifies the degree of unpredictability of a random variable's possible outcomes:

$$H := -\sum_{k=-\infty}^{\infty} p_k \log_2 p_k \,. \tag{4}$$

2.2 Working Hypothesis

Our working hypothesis states that **the ℓ_1-norm is the natural measure over the integer lattice**, since it directly measures the sum of absolute differences along each coordinate axis. At the same time, the ℓ_2-norm, which considers the Euclidean interpretation, provides a distorted information and may thus introduce imprecision when analyzing data on discrete lattices – hence becoming an unnatural measure. Revisiting Figure 1 to justify this working hypothesis, points that are placed on the ℓ_1-sphere maintain a consistent Hamming distance. In contrast, for points on an ℓ_2 sphere, this direct relationship breaks down. The same ℓ_2 distance from the origin does not guarantee a consistent Hamming distance, which can complicate analyses relying on both metrics. We therefore adopt the ℓ_1-norm and claim that the shift from the ℓ_2-norm within ESs' existing work necessitates careful consideration of basic components such as the mutative step-size, but it also impacts other considerations, including statistical measures. Importantly, the *statistical covariance matrix* is interpretable by means of the ℓ_2-norm: the variance of a single variable is the square of the ℓ_2-norm of its deviation from the mean, and the covariance between two variables is the ℓ_2-norm of their joint deviations. As such, the covariance matrix is not necessarily a proper measure to consider in the context of correlations among variables over the integer lattice. The ℓ_2-based Pearson correlation measure also falls into this category, and therefore has a limited potential to statistically describe such variables.

2.3 Discrete Mutation Distributions

Studying the effectiveness of discrete randomized heuristic search in integer search-spaces dates back to the 1960's and 1970's. Kelahan and Gaddy proposed to use a bilateral power distribution that adaptively shrank by a geometric schedule [14]. Rudolph's early research, dating back to 1994, questioned the suitability of random distributions for evolutionary mutations in unbounded integer spaces [22]. In particular, Rudolph showed that integer mutations utilizing the geometric distribution possess maximum entropy, and demonstrated the effectiveness of such mutations in practice. The goal of this section is to present in detail two probability distributions of interest, the TN and DG distributions. We begin by mentioning two additional discrete distributions, which are intuitive and may serve as reference distributions.

2.3.1 The Discrete Uniform (DU) Distribution. Playing the role of the most intuitive manner to generate random integers, the discrete uniform distribution of interest draws a random variable from a predefined range at equal probabilities, which in our context of search mutations corresponds to $\{-N,\ -N+1, \ldots,\ 0, \ldots,\ N-1,\ N\}$:

$$\mathbf{Pr}\{X = k\} = \frac{1}{2N+1} \quad \forall k \in \{-N, -N+1, \ldots, N-1, N\}\,. \tag{5}$$

This range dictates the variation's strength, and therefore, it is possible to control the mutation step by considering the expected ℓ_1-norm, which reads $S_{DU} = n \cdot \frac{N(N+1)}{2N+1}$. Given a desired mean step-size S_*, by solving for N, it may be set to

$$N_* := \texttt{round}\left[\frac{2(\frac{S_*}{n}) - 1 + \sqrt{1 + (\frac{S_*}{n})^2}}{2}\right]. \tag{6}$$

The entropy of this distribution over the finite support range is

$$H_{DU} = \log_2(2N+1), \tag{7}$$

which constitutes the maximal entropy function among all discrete distributions over the same finite support range. Our investigation of the entropy functions in Section 4 will consider the unbounded integer space and an accompanying constraint in the form of a bounded variance. Nevertheless, this DU distribution will serve as an important reference.

2.3.2 The Shifted Binomial (SB) Distribution. Describing a certain number of successful trials out of a fixed budget N, we consider this discrete probability distribution in our integer context with an equal probability for success/failure ($p = \frac{1}{2}$). Focusing on the symmetric case which enables generation of negative as well as positive values, the random variable is shifted, $Y := X - \frac{N}{2}$, where X is the binomial random variable with parameters N and $p = \frac{1}{2}$. The probability of Y to get a value k is given by

$$\mathbf{Pr}\{Y = k\} = \mathbf{Pr}\left\{X = k + \frac{N}{2}\right\} = \binom{N}{k + \frac{N}{2}} p^{k+\frac{N}{2}} (1-p)^{N-(k+\frac{N}{2})}. \tag{8}$$

This distribution holds a significant place in the history of ESs, a role that stemmed from the practical use of *Galton boards* to generate mutations. Rechenberg notably exercised this approach during his 1964 experimental optimization campaign, aimed at minimizing drag in a kinked plate setup [18, p. 26]. Notably, the budget of trials N defines the range of generated numbers, and in our context of integer mutations – it governs the step's strength. Equivalently to the DU distribution, it is possible to control the step by considering the expected ℓ_1-norm. For this distribution, it may be approximated via the so-called Mean Absolute Deviation which yields $S_{SB} \approx n \cdot \sqrt{\frac{2N}{\pi}}$. Accordingly, given a desired step-size S_*, the distribution can be tuned accordingly by computing N_*:

$$N_* := \texttt{round}\left[\frac{\pi}{2}\left(\frac{S_*}{n}\right)^2\right]. \tag{9}$$

While offering an artificial means of step-size control, this distribution's lack of an infinite tail inherently limits its natural fit for unbounded search problems, which ideally require the capacity for arbitrarily large mutations to explore the full solution landscape.

2.3.3 The Truncated Normal (TN) Distribution. The Normal distribution constitutes the continuous limit of the Binomial distribution when the trials' number n increases and becomes sufficiently large. Given a normally distributed random variable, $z \sim \mathcal{N}(0, \sigma^2)$, which may take any value within $[-\infty, \infty]$, it is rounded and then denoted as $\texttt{round}(z)$. The probability of this rounded z to take to an integer value of k is given by:

$$\begin{aligned}\mathbf{Pr}\{\texttt{round}(z) = k\} &= \mathbf{Pr}\{k - 0.5 \le z < k + 0.5\} \\ &= \Phi\left(\frac{k+0.5}{\sigma}\right) - \Phi\left(\frac{k-0.5}{\sigma}\right),\end{aligned}$$

where $\Phi(x)$ is the cumulative distribution function (CDF) of the standard normal distribution, $\Phi(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$, with $\mathrm{erf}(x)$ being the error function defined as $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,\mathrm{d}t$. The explicit probability thus becomes

$$\boxed{\mathbf{Pr}\{\texttt{round}(z) = k\} = \frac{1}{2}\left[\mathrm{erf}\left(\frac{k+0.5}{\sqrt{2}\sigma}\right) - \mathrm{erf}\left(\frac{k-0.5}{\sqrt{2}\sigma}\right)\right].} \tag{10}$$

This term can be approximated using the Mean Value Theorem (MVT) for the error function, or simply by taking the first-order Taylor expansion (denote the function's argument as $x_\pm := \frac{k \pm 0.5}{\sqrt{2}\sigma}$):

$$\mathrm{erf}(x_+) - \mathrm{erf}(x_-) \approx \mathrm{erf}'(\bar{x}) \cdot (x_+ - x_-) \approx \frac{2}{\sqrt{\pi}} \exp\left(-\bar{x}^2\right)$$

where $\bar{x} = \frac{k}{\sqrt{2}\sigma}$ serves as the mean value.
This approximation holds for small $\bar{x}$ values, which translates to moderate to large standard deviations of the normal distribution (the regime where $\sigma \gtrsim 0.75$). Otherwise, the regime of smaller σ values induces a probability of nearly one for no mutations, i.e., $\sigma < 0.75 \rightsquigarrow \mathbf{Pr}\{\texttt{round}(z) = 0\} \approx 1.0$.[2] We are now in a position to **approximate** the probability function of (10) via the following:

$$\boxed{\mathbf{Pr}\{\texttt{round}(z) = k\} \approx \frac{1}{\sqrt{\pi}} \exp\left(-\frac{k^2}{2\sigma^2}\right).} \tag{11}$$

Next, we would like to express the expected ℓ_1-norm of this TN distribution. Attempts to calculate it in a compact form have been *unsuccessful*, even upon using the approximated density function,

$$S_{TN} := n \cdot \mathbb{E}\left[|\texttt{round}(z)|\right] \approx \frac{n}{\sqrt{\pi}} \sum_{k=-\infty}^{+\infty} |k| \cdot \exp\left(-\frac{k^2}{2\sigma^2}\right). \tag{12}$$

We hypothesize that the expectation value of the continuous Normal distribution serves as a fine approximation for the discrete counterpart. Given the expected ℓ_1-norm as calculated exactly for the standard Normal distribution with zero-mean,

$$\mathbb{E}\left[\|\vec{z}\|_1\right] = n \cdot \int_{-\infty}^{\infty} |z| \cdot \mathrm{pdf}(z)\,\mathrm{d}z = n\sigma\sqrt{\frac{2}{\pi}} \tag{13}$$

we empirically validated it for the TN distribution over a large range of σ values across various dimensions. Figure 2 presents these

[2] Another decent approximation for $\mathrm{erf}(x)$ is the *hyperbolic tangent* function,

$$\mathrm{erf}(x) \approx \mathrm{tgh}(x) := \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)},$$

by which the difference can be presented as

$$\begin{aligned}\mathrm{erf}(x_+) - \mathrm{erf}(x_-) \approx& \frac{2\left(\exp\{x_+ - x_-\} - \exp\{-x_+ + x_-\}\right)}{\left(\exp\{x_+\} + \exp\{-x_+\}\right)\left(\exp\{x_-\} + \exp\{-x_-\}\right)} \\ =& \frac{2\left(\exp\{\frac{1}{\sqrt{2}\sigma}\} - \exp\{\frac{-1}{\sqrt{2}\sigma}\}\right)}{\exp\frac{\sqrt{2}k}{\sigma} + \exp\frac{1}{\sqrt{2}\sigma} + \exp\frac{-1}{\sqrt{2}\sigma} + \exp\frac{-\sqrt{2}k}{\sigma}}\end{aligned}$$

without leading to a concise representation yet yielding similar numerical patterns.

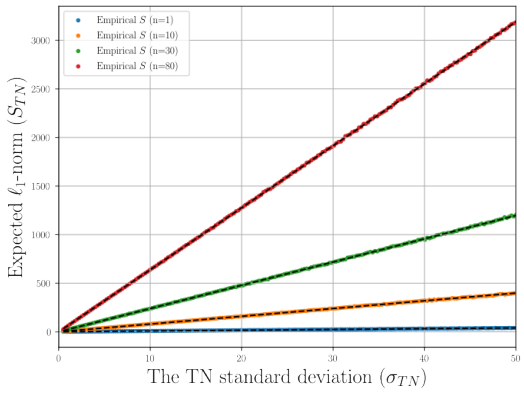

Figure 2: Expected l_1-norm S vs. σ for the TN Distribution per various vector dimensions: $n \in \{1, 10, 30, 80\}$. The curves depict the empirical expectation (via colored circles) while exercising the approximation of S_{TN} given in Eq. (14) (via a dashed line) – which demonstrate excellent fit.

empirical validation tests for $n = \{1, 10, 30, 80\}$, where evidently this approximation's accuracy improves as the dimensionality increases. We conclude that this relation, i.e.,

$$S_{TN} \approx n\sqrt{\frac{2}{\pi}} \cdot \sigma_{TN} \quad \Longleftrightarrow \quad \sigma_{TN} \approx \sqrt{\frac{\pi}{2}} \cdot S_{TN}/n, \tag{14}$$

serves as a fine approximation per the TN distribution for the expected ℓ_1-norm, even for small σ values.

Finally, we denote the entropy function of this distribution using the exact probability of Eq. (10) as H_{TN}, and also specify the approximated entropy using Eq. (11):

$$\tilde{H}_{TN} \approx -\sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{\pi}} \exp\left(-\frac{k^2}{2\sigma^2}\right) \cdot \log_2\left(\frac{1}{\sqrt{\pi}} \exp\left(-\frac{k^2}{2\sigma^2}\right)\right). \tag{15}$$

2.3.4 The Double Geometric (DG) Distribution. Rudolph's study proposed to use the doubly-geometric mutation, which possesses a symmetric distribution with respect to 0. It is achieved by drawing two random variables, $\{g_1, g_2\}$, according to the geometric distribution, $\mathbf{Pr}\{g_J = k\} = p \cdot (1-p)^k$ ($J = 1, 2$), and taking their difference, $z = g_1 - g_2$. The probability function of z reads,

$$\boxed{\mathbf{Pr}\{z = k\} = \frac{p}{2-p} \cdot (1-p)^{|k|}, \quad k \in \mathbb{Z}} \tag{16}$$

with $\mathbb{E}[z] = 0$ and $\mathrm{VAR}[z] = 2(1-p)/p^2$. Next, to generalize to the multivariate case of an n-dimensional mutation vector $\vec{z}$, Rudolph showed that by taking n stochastically independent random variables, following the doubly-geometric distribution, the properties of symmetry and maximal entropy are kept. **In practice**, each random variable is drawn by the following calculation (g_J are *geometrically* distributed random variables, both with parameter p):

$$\begin{aligned} g_J &\longleftarrow \left\lfloor \frac{\log(1 - \mathcal{U}(0,1))}{\log(1-p)} \right\rfloor \quad J = 1, 2 \\ \mathcal{G}_n(0, p) &:= g_1 - g_2. \end{aligned} \tag{17}$$

Importantly, this multivariate distribution could be *controlled by the mean step-size* $S = \mathbb{E}[\|\vec{z}\|_1] = n \cdot \mathbb{E}[|z_1|]$,

$$S_{DG}(p) = n \cdot \frac{2(1-p)}{p(2-p)} \quad \Longleftrightarrow \quad p = 1 - \frac{S_{DG}/n}{\sqrt{\left(1 + (S_{DG}/n)^2\right)} + 1}. \tag{18}$$

Finally, we specify the entropy function of the DG distribution:

$$H_{DG} = -\sum_{k=-\infty}^{\infty} \frac{p}{2-p} \cdot (1-p)^{|k|} \cdot \log_2\left(\frac{p}{2-p} \cdot (1-p)^{|k|}\right). \tag{19}$$

2.3.5 Numerical Assessment. We choose to focus on the TN and DG distributions and numerically assess the accuracy of the aforementioned formulations. We empirically compute the mean ℓ_1- and ℓ_2-norm of randomly generated n-dimensional integer vectors. To introduce variability, we generate those vectors using an increasing scale of step-sizes along their coordinates,

$$S_i = K \cdot i \quad \forall i \in 1 \dots n,$$

and systematically inflate it along a spectrum of factors $K \in \{1, \dots, 50\}$ across multiple dimensions $n \in \{2, 10, 30, 80\}$. This numerical evaluation is presented as Figure 3, with the theoretical expected ℓ_1 step-size being $S = K \cdot \sum_i i = \frac{1}{2}Kn(n+1)$ due to Eq. (3). The results corroborate the validity of our formulations per the ℓ_1-norm, while showing a good fit between the theoretical and the empirical mean values, scalable across K and n. At the same time, **this numerical evaluation demonstrates the inability of the ℓ_2-norm to quantify the integer vectors over the discrete lattice**, becoming unreliable when $n \gtrsim 30$.

Next, by setting S as the driving step-size, and by exercising the relations $\sigma_{TN}(S)$ and $p_{DG}(S)$ (using Eqs. (14) and (18), respectively), we compare the two distributions over a set of S values and present their histograms as a gallery in Figure 4. Evidently, their shapes differ – while the TN exhibits relative smoothness and strong resemblance to the convex Bell curve as expected, the DG distribution possesses a spiked, concave shape with a pronounced peak at zero.

3 CORRELATED SAMPLES VIA ROTATIONS

The challenge of well-defining rotations over the integer lattice stems from the fundamental incompatibility between the geometric concept of rotation, which involves trigonometric functions that induce irrational coordinate values, and the discrete, axis-aligned structure of the lattice. An ambition to define ℓ_1-norm-preserving rotations is unrealistic in the generalized case, and can be fulfilled only for a limited set of special cases:

(i) Permutation of coordinates, e.g., in the 2D case via the transformation

$$R_{\text{perm}} := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

(ii) Sign changes of coordinates, e.g., in the 2D case via the transformation

$$R_{\text{sign}} := \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

(iii) Combinations of the above (i)+(ii).

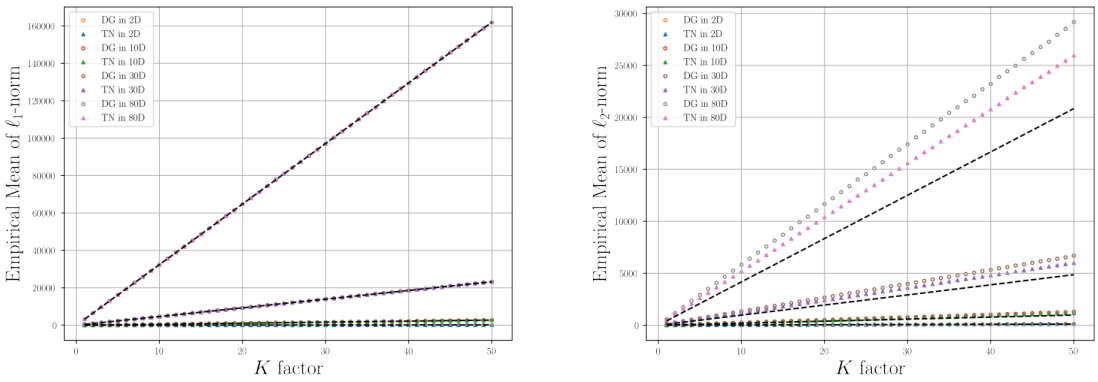

Figure 3: Numerically assessing the DG and TN distributions: The empirical mean of the ℓ_1- versus the ℓ_2-norm of populations of randomly generated n-dimensional integer vectors with an increasing scale of individual step-sizes, that is $S_i = K \cdot i$, subject to a factor $K \in \{1, \ldots, 50\}$ across multiple dimensions $n \in \{2, 10, 30, 80\}$. The theoretical step-sizes are depicted as the black dashed lines. [LEFT] ℓ_1-norm, and [RIGHT] ℓ_2-norm. Evidently, the ℓ_2-norm becomes unreliable for quantifying vectors when $n \gtrsim 30$.

We propose a rotation transformation, noting that similar procedures for continuous distributions exist in the literature (e.g., [11]).

3.1 Proposed Rotation Transformation

Dropping the requirement for ℓ_1-norm preservation, given an uncorrelated mutation instance $\vec{z}_u$, a correlated mutation instance $\vec{z}_c$ is attainable by *rounding to the nearest integer* the sequence of $n(n-1)/2$ rotations using the operator $\mathbf{R}(\theta) := (r_{k\ell})$

$$\vec{z}_c = \texttt{round}\left[\left(\prod_{i=1}^{n-1}\prod_{j=i+1}^{n}\mathbf{R}(\alpha_{ij})\right)\cdot\vec{z}_u\right]. \tag{20}$$

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

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

where the angle $r_{k\ell}$ represents a correlation measure between the k^{th} and ℓ^{th} random variables. Overall, this procedure is well-defined when given an uncorrelated mutation vector $\vec{z}$ and a vector of rotational angles $\vec{\alpha}$:

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

The Original Continuous Usage. A "pure", rounding-free transformation, i.e., $\vec{z}_c = \left(\prod_{i=1}^{n-1}\prod_{j=i+1}^{n}\mathbf{R}(\alpha_{ij})\right)\cdot\vec{z}_u$, played the rotation role for generating correlated normal mutations in Schwefel's formulation of the Standard ES [1, 24]. This transformation was designed to store covariance information by means of the n-dimensional variances' vector $\vec{\sigma}$ as well as the $n(n-1)/2$-dimensional vector of rotational angles $\vec{\alpha}$. In the continuous domain, the statistical covariance of the decision variables is meaningful, also due to its hypothesized relation to the inverse Hessian matrix of the search landscape (which was later confirmed [28]). Therefore, it was important to design transformations that account for such information and store it. The transformation of a covariance element c_{ij} into a rotational angle α_{ij} (where $c_{ii} \equiv \sigma_i^2$) provides another useful relationship for decision variables i and j:

$$\alpha_{ij} = \frac{1}{2}\arctan\left(\frac{2c_{ij}}{\sigma_i^2 - \sigma_j^2}\right), \tag{21}$$

where $\alpha_{ij} = 0$ whenever no correlation exists. Indeed, this representation of variances and angles was preferred in the Standard ES over the canonical matrix form (c_{ij}) since it was easier to maintain the mathematical property of the covariance being positive-definite when the update occurs via self-adaptation. Later on, Rudolph [21] verified the validity of this representation and showed that the usage of such angles does not restrict the generality of correlated ES mutations.

In our context of integer search, we can use the aforementioned `rotateInt` procedure to generate correlated integer mutations. However, our usage is not concerned by any mathematical constraint to form a proper covariance matrix, whose interpretation is limited and usefulness unknown over the integer lattice.
Such a procedure is proposed as Algorithm 1, relying on the `genUncorrelatedMutation()` subroutine (see Supplementary Material), which generates the uncorrelated samples by their type (denoted as {`DG`, `TN`}; note that a single call of `DG` obtains a geometric sample, and thus it is needed to call it twice and take the difference).

We will numerically evaluate this procedure in the next subsection. Notably, it is possible to artificially form a matrix using a step-size vector $\vec{S}$ and rotational angles $\vec{\alpha}$ by exercising Eqs. (14) and (21), respectively. However, this construction is not guaranteed to hold the necessary covariance matrix mathematical properties, with positive semidefiniteness (PSD) being the primary.
Whenever it is mathematically sound, after applying a numerical procedure to enforce the PSD property, we will consider such a constructed covariance matrix and denote it by $\mathbf{C}_{FTN}$ (which

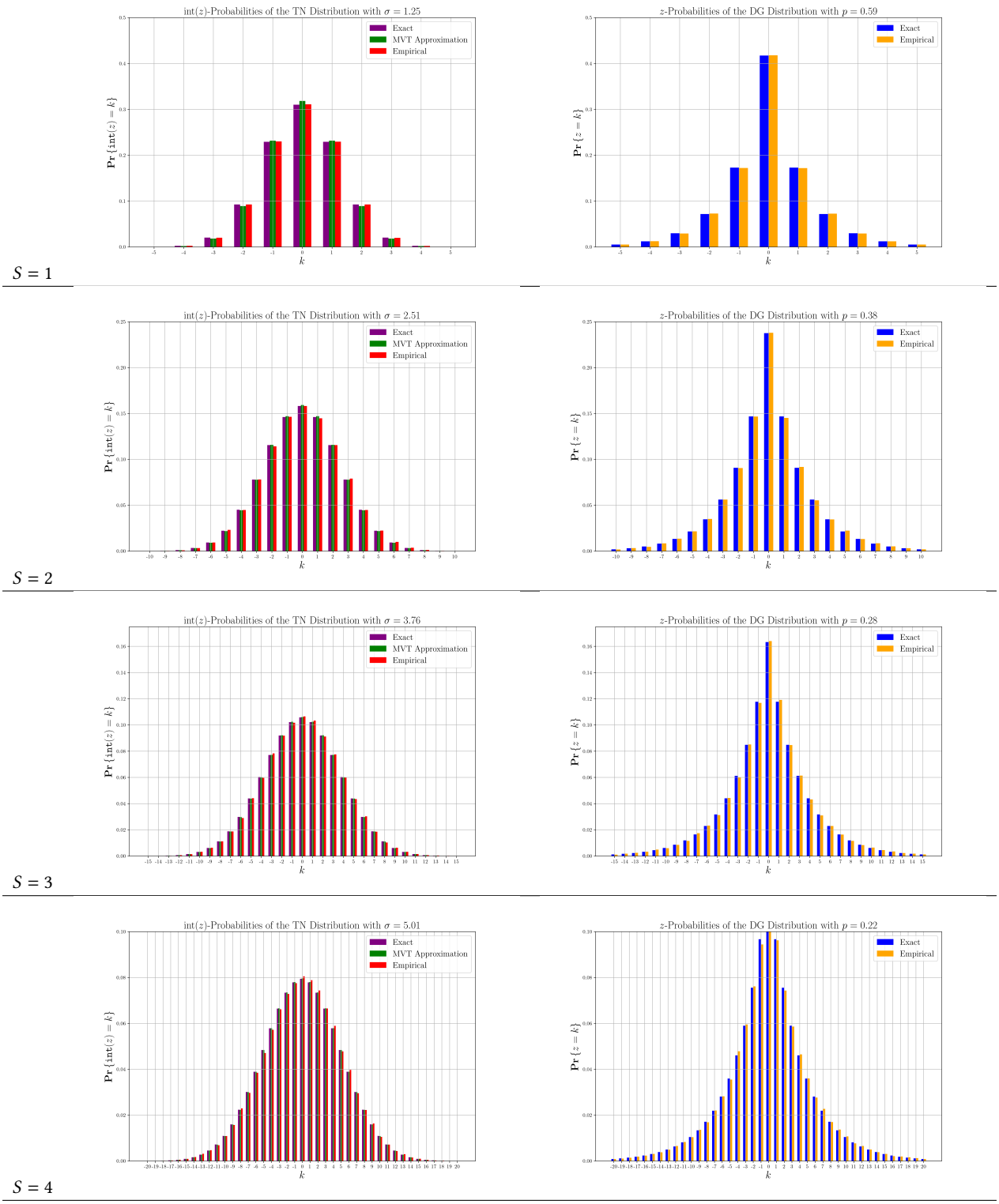

Figure 4: The single-variable TN and DG distributions per four concrete mean step-size values $S \in \{1, 2, 3, 4\}$ (top to bottom): [LEFT] TN probabilities (exact versus MVT approximation versus empirical) using $\sigma_{TN}(S)$ from Eq. (14) (explicit σ values are specified on each plot); [RIGHT] DG probabilities (exact versus empirical) per the computed p when accounting for S_{DG} using (18) (explicit p values are specified on each plot).

corrMutate$(\vec{S},\ \vec{\alpha},\ n,\ \texttt{type})$
  $\vec{z}_u \longleftarrow \texttt{genUncorrelatedMutation}\left(\vec{S},\ \texttt{type}\right)$
  $\vec{z} \longleftarrow \texttt{rotateInt}\left(\vec{z}_u,\ \vec{\alpha}\right)$
  if type==DG **then**
    $\vec{z}_g \longleftarrow \texttt{genUncorrelatedMutation}\left(\vec{S},\ \texttt{type}\right)$
    $\vec{z}'_g \longleftarrow \texttt{rotateInt}\left(\vec{z}_g,\ \vec{\alpha}\right)$
    $\vec{z} \longleftarrow \vec{z} - \vec{z}'_g$ // difference of two geometric samples
  end
  return $\{\vec{z}\}$

Algorithm 1: A procedure to generate a correlated mutation vector of integers. $\left\{\vec{S}, \vec{\alpha}\right\}$ are the step-sizes and rotation angles, respectively. This procedure handles either the DG or the TN distributions via its subroutine `genUncorrelatedMutation()`, which accounts for the type (denoted as {DG, TN}). Calling this subroutine with DG obtains a single geometric sample, and hence the need to call it twice and take the difference. Finally, the rotation, via `rotateInt()`, is agnostic with respect to the underlying distribution of its uncorrelated input sample.

stands for Forced TN). Then, we can use it to generate correlated multivariate normally distributed vectors with truncation:

$$\vec{z}_{FTN} \sim \texttt{round}\left[\mathcal{N}\left(\vec{0}, \mathrm{C}_{FTN}\right)\right] . \tag{22}$$

3.2 Numerical Assessment of 2D Rotations

Heatmaps. To visually demonstrate the ability to generate correlated integer mutations, we present populations of 2D samples per the TN and DG distributions across multiple rotation angles with $S_1 = 1,\ S_2 = 2$ and $\theta \in \{0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{\pi}{2}\}$ – via heatmaps (see Figure 5). The DG distribution exhibits a spiked pattern, in contrast to the smooth TN pattern, where mind should be given to the scale. This pattern constitutes a two-dimensional extension of the single variable case, which also demonstrated discrete versus smooth profiles, respectively (Figure 4).

ℓ_1-norm under systematic rotations. Next, we randomly generated populations of 2D integer vectors using the two distributions when subject to *systematic* rotation by continuous angle $\theta \in [0, \frac{\pi}{2}]$ under four settings $S_2 \in \{2, 3, 4, 5\}$ with $S_1 = 1$. We empirically compute the mean ℓ_1-norm of such rotations in Figure 6, where the FTN samples are partially defined over this range of θ. It is evident that the proposed rotation does not preserve the ℓ_1-norm, which is to be expected, and that the bias is maximized at the $\theta = \frac{\pi}{4}$ rotational angle. Also, the DG distribution consistently obtains higher ℓ_1-norm values with a less smooth pattern when compared to the TN distribution.

Statistical measures for correlations. Finally, we tested four statistical measures to quantify the correlation of the two integer decision variables in the 2D case. We generated the populations similarly to the previous simulation of the ℓ_1-norm under rotations. The Pearson correlation measure failed to provide a fair representation of the data and is excluded from this report. Figure 7 presents two correlation measures over randomly generated 2D integer vectors per the TN/DG distributions when subject to systematic rotation by $\theta \in [0, \frac{\pi}{2}]$ under four settings $S_2 \in \{2, 3, 4, 5\}$ with $S_1 = 1$: the default statistical covariance, and the covariance over absolute values (that is, empirical covariance over $|z_i|$ values). We also computed ℓ_1-driven correlation via the LASSO coefficients [29] (see Supplementary Material). The FTN samples (depicted by black points in both figures) are defined only partially over this range of θ. The DG distribution consistently exhibits abrupt changes in its calculated measures, but its trend curves are consistent.

Setting aside the FTN data points (which are anyway not well-defined), it is evident that the three statistical measures are able to quantify the correlation proportionally to the rotation in an expected pattern. Assessing the effectiveness of the considered statistical measures in meaningfully expressing correlations driven by the ℓ_1-norm requires further research and is not addressed here.

4 ENTROPY

Rudolph investigated the family of maximum entropy **discrete** distributions [22] by addressing the following optimization problem:

$$\begin{aligned} \text{maximize}_{p_k} \quad & H := -\sum_{k=-\infty}^{\infty} p_k \log p_k \\ \text{subject to:} \quad & p_k = p_{-k}, \quad \forall k \in \mathbb{Z}, \\ & \sum_{k=-\infty}^{\infty} p_k = 1, \\ & \sum_{k=-\infty}^{\infty} k^2 p_k = \sigma^2, \\ & p_k \geq 0, \quad \forall k \in \mathbb{Z}. \end{aligned} \tag{23}$$

Entropy of statistical distributions is typically measured under a given constraint – here, Rudolph set this constraint to be a *bounded variance* (linked to the desired step-size), arguing that otherwise the entropy analysis becomes an underdetermined problem. By treating the problem's **Lagrangian** $\mathcal{L}(p, a, b)$ (neglecting the last condition)

$$\mathcal{L} = -\sum_{k=-\infty}^{\infty} p_k \log p_k + a\left(\sum_{k=-\infty}^{\infty} p_k - 1\right) + b\left(\sum_{k=-\infty}^{\infty} k^2 p_k - \sigma^2\right) \tag{24}$$

Rudolph concluded that the DG distribution constitutes the *maximizer* of this optimization problem. Next, we would like to assess the entropy functions of the considered distributions.

4.1 Entropy versus Mean Step-Size (1D)

In order to establish a baseline for comparing the distributions, a common mutation step-size must be well-defined. We are able to draw a comparison of the entropy as a function of the integer mutation step-size, by considering the mean step-size S as the driving parameter to dictate σ_{TN} and p. Accordingly, we compare the entropy functions that are derived from the two distributions – H_{TN} and H_{DG} – and additionally consider H_{SB} (using its exact probability function (8)) by exercising (9) as well as H_{DU} (7).
Figure 8 depicts the entropy functions over a spectrum of S, using the specified relations. While the DU distribution exhibits the

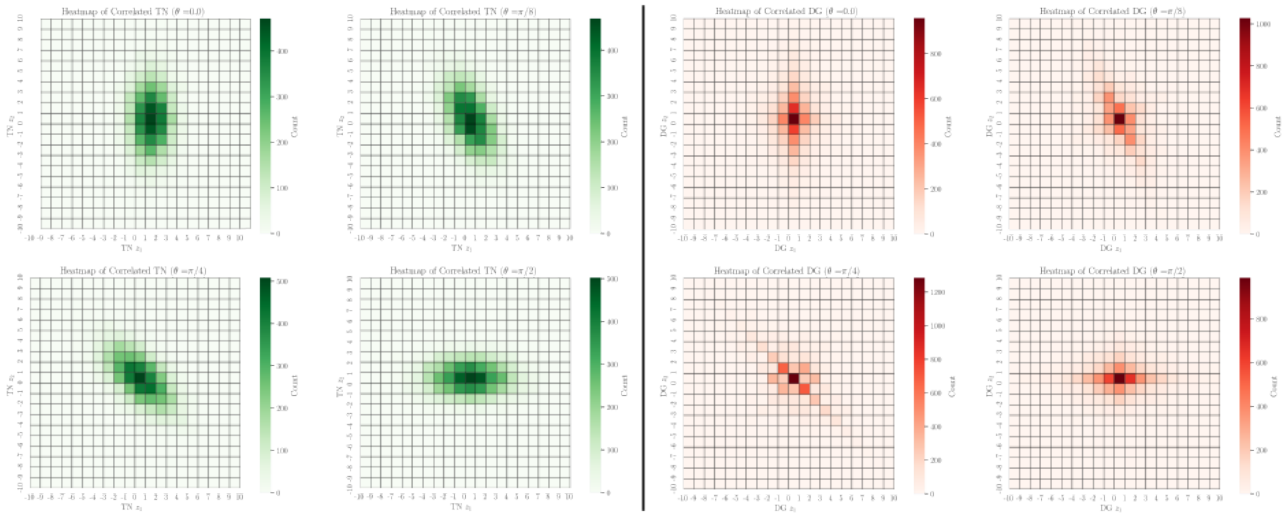

Figure 5: Heatmaps depicting populations of rotated 2D samples with $S_1 = 1$, $S_2 = 2$ and $\theta \in \{0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{\pi}{2}\}$ (set in this order clockwise - see titles). [LEFT, green]: TN distribution, [RIGHT, red]: DG distribution. The population size was set to 10^4.

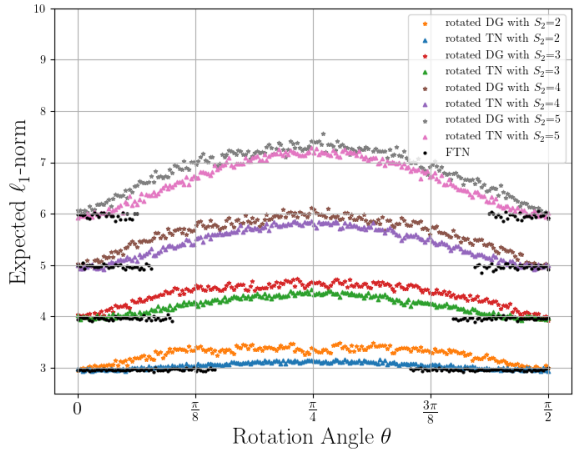

Figure 6: The empirical ℓ_1-norm of randomly generated 2D integer vectors per the TN/DG distributions when subject to systematic rotation with $\theta \in [0, \frac{\pi}{2}]$ under four settings $S_2 \in \{2, 3, 4, 5\}$ with $S_1 = 1$. The population size was set to 10^4.

lowest entropy values, the entropy functions of the SB and TN distributions practically merge. Most importantly, it is evident that the DG distribution consistently possesses the highest entropy values, constituting a numerical validation to Rudolph's result.

4.2 Estimated Entropy of 2D Samples

In order to numerically assess the entropy values of 2D randomly generated integer vectors, we turn to histogram-based estimations of the generated populations. We consider 2D populations drawn from the DG and TN distributions when controlled by S_1 and S_2. Firstly, uncorrelated samples are generated with $S_1 = 1$ over a spectrum of S_2 values, and their estimated entropy values are presented in Figure 9[top]. Secondly, correlated samples subject to $S_1 = 1$ and $S_2 \in \{2, 3, 4\}$ are generated over a spectrum of rotation angles $\theta \in [0, \pi/2]$ and undergo histogram estimation to obtain the entropy values – depicted in Figure 9[bottom].

Evidently, the samples generated by the DG distribution consistently possess higher entropy values in all the investigated 2D scenarios – suggesting that Rudolph's result holds also under the proposed rotation transformation.

5 PRACTICAL OBSERVATIONS

We present empirical observations of IES variants across quadratic models and discuss them in light of the aforementioned theoretical principles. These observations are based on a recent empirical study that investigated the performance of IESs over integer quadratic programming (IQP) problems [25]. The IES is fully described in the Supplementary Material.

5.1 Standard IES

By adopting the proposed correlated mutation operator (`corrMutate` of Algorithm 1), it is necessary to define the self-adaptation scheme of the strategy parameters $\vec{S}$ and $\vec{\alpha}$. Adhering to Schwefel's original scheme, the update steps read

$$S_i' \longleftarrow S_i \cdot \exp\left\{\tau_g \cdot \mathcal{N}_g + \tau_\ell \cdot \mathcal{N}_i(0,1)\right\} \quad \text{for } i = 1 \ldots n\,, \tag{25}$$

$$\alpha_j' \longleftarrow \alpha_j + \beta \cdot \mathcal{N}_j(0,1) \quad \text{for } j = 1, \ldots, n \cdot (n-1)/2\,, \tag{26}$$

using the default parametric settings: $\tau_g := \frac{1}{\sqrt{2 \cdot n}}$, $\tau_\ell := \frac{1}{\sqrt{2 \cdot \sqrt{n}}}$ and $\beta := 0.0873$ rad [1]. Additionally, the renowned 1/5th success-rule [1] for the step-size adaptation is considered in play with either the TN or DG mutation distributions. Altogether, six IESs variants were considered, reflecting the combinations across the underlying distributions DG/TN and the application of rotations: (1+1)-DG, (1+1)-TN, correlatedDG, correlatedTN, uncorrelatedDG,

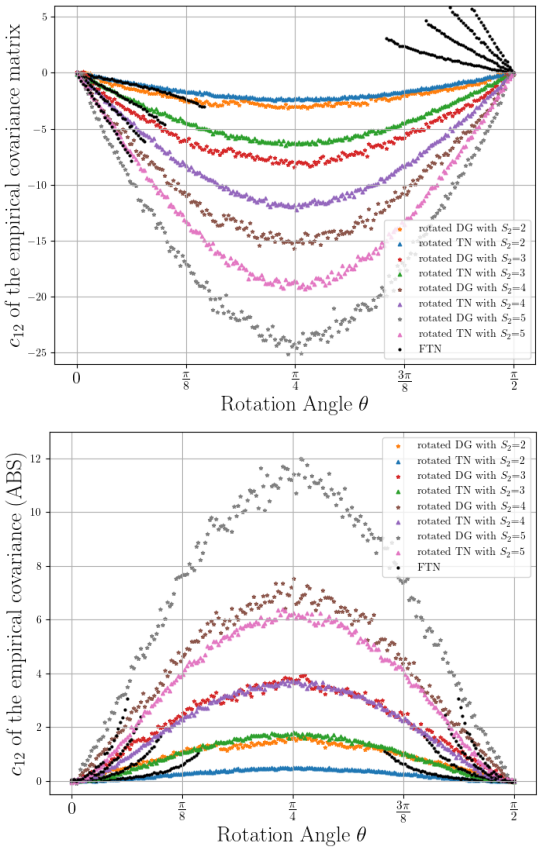

Figure 7: Numerical assessment of various correlation measures over randomly generated 2D integer vectors per the TN/DG distributions when subject to systematic rotation with $\theta \in [0, \frac{\pi}{2}]$ under four settings $S_2 \in \{2, 3, 4, 5\}$ with $S_1 = 1$: (LEFT) default statistical covariance, and (RIGHT) covariance over absolute values ($|z_i|$). The FTN samples (depicted by black points) are defined only partially over this range.

and uncorrelatedTN. The DU distribution has been tested and consistently failed to deliver competitive results (numerical results are excluded), implying that the distribution plays an important role. Analyses have been carried out by the `IOHanalyzer` tool [31].

5.2 Preliminary: Stagnation over the Sphere

The six IESs were tested over the integer Sphere function by means of 50 runs over three dimensionalities $n \in \{30, 50, 80\}$. The summary of the 80D runs via the fixed-budget perspective is provided in Figure 10 as an Expected Target Value plot [31]. The (1+1)-DG clearly outperforms the other strategies on this test-case, but it is also evident that all six IESs do not reach the optimum in most of the runs. Instead, they enter stagnation phases away from the optimum, even when no local optima exist. This behavior is consistently observed across other dimensionalities. Thus, we argue that this stagnation stems from inherent issues within the IESs' underlying mechanisms, independent of the search landscape's difficulty.

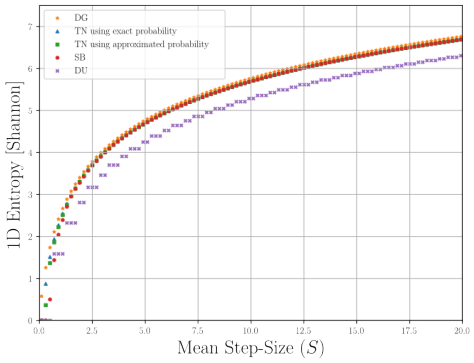

Figure 8: The entropy function of the single-variable distributions over the spectrum of S, which controls their defining step-size: H_{TN} and $\tilde{H}_{TN}$, via $\sigma_{TN}(S)$ from Eq. (14), using the exact (10) and the approximated (11) probabilities, respectively – which yield strong resemblance; H_{DG} (19) over p (via the relation $S_{DG}(p)$ (17) when accounting for $p \in [0, 1]$), and additionally H_{SB}, using its exact probability function (8) and by exercising (9) – showing a strong resemblance to the TN distribution, as expected from the Binomial-Normal relation.

5.3 Integer Quadratic Programming (IQP)

The four main IESs variants, excluding the (1+1) variants, were applied to IQP problems of the form:

$$\begin{aligned} &\text{minimize}_{\vec{x}} \quad \tfrac{1}{c} \cdot \left[\left(\vec{x} - \vec{\xi}_0\right)^T \mathbf{H} \left(\vec{x} - \vec{\xi}_0\right) \right] \\ &\text{subject to:} \quad \vec{x} \in \mathbb{Z}^n. \end{aligned} \tag{27}$$

where the Hessian matrix $\mathbf{H}$, its parametric condition number c and the location vector $\vec{\xi}_0$ completely define a problem instance. To instantiate a testbed, the commonly used Ellipsoidal, Discus and Cigar functions [10][3] were considered via four Hessian matrices:

- **H-1** Discus: $(\mathcal{H}_{\text{disc}})_{11} = c$, $(\mathcal{H}_{\text{disc}})_{ii} = 1 \quad i = 2, \ldots, n$;
- **H-2** Cigar: $\left(\mathcal{H}_{\text{cigar}}\right)_{11} = 1$, $\left(\mathcal{H}_{\text{cigar}}\right)_{ii} = c \quad i = 2, \ldots, n$;
- **H-3** Rotated Ellipse: $\mathcal{H}_{\text{RE}} = \mathcal{O}\mathcal{H}_{\text{ellipse}}\mathcal{O}^{-1}$, 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: $\mathcal{H}_{\text{HE}} = \mathcal{S}\mathcal{H}_{\text{ellipse}}\mathcal{S}^{-1}$, where the rotation constitutes the normalized Hadamard matrix, $\mathcal{S} := \text{Hadamard}(n)/\sqrt{n}$.

providing two separable and two non-separable scalable test-cases. Six levels of conditioning, $c \in \left\{10, 10^2, \ldots, 10^6\right\}$ yielded altogether 24 problem instances per dimensionality. Systematic benchmarking of the IESs over this testbed included three search-space dimensionalities $n \in \{32, 64, 128\}$. Figure 11 presents pairwise comparisons of the four IESs via heatmaps, according to the fraction of times their mean values are better over all the functions considered, per $n = 64$, where the calculations were conducted in isolation for both the separable (left) and non-separable subsets (right). This behavioral

[3]These functions correspond to $\{f_2, f_{10}, f_{11}, f_{12}\}$ at the renowned BBOB suite. The actual rotation of f_{10} is implemented as reported in [28].

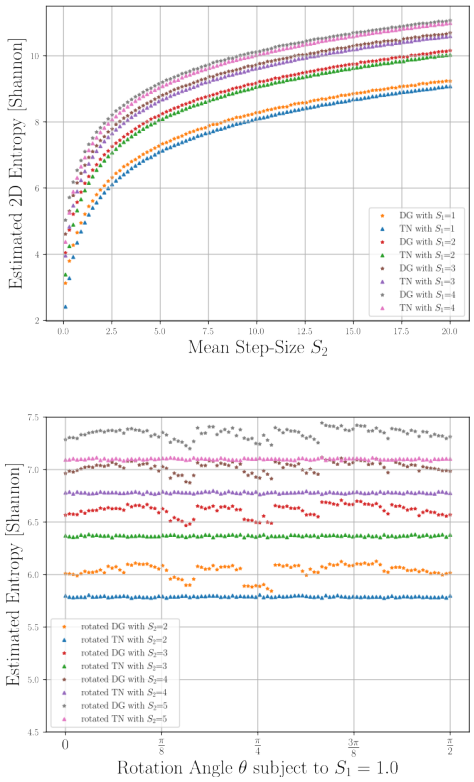

Figure 9: Entropy histogram-based estimations of the 2D DG and TN distributions when controlled by S_1 and S_2. Uncorrelated samples with multiple step-sizes: $S_1 \in \{1, 2, 3, 4\}$ over a spectrum of S_2 values (top), and correlated samples subject to $S_1 = 1$ and $S_2 \in \{2, 3, 4, 5\}$ over a spectrum of rotation angles $\theta \in [0, \pi/2]$ (bottom). Evidently, the DG distribution consistently possesses higher entropy values in all the 2D scenarios, both correlated and uncorrelated.

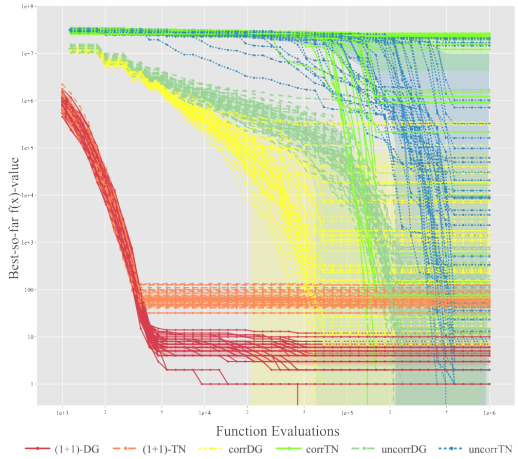

Figure 10: The six IESs variants applied to the 80D integer Sphere function: the Expected Target Value plot of the fixed-budget perspective over the individual 50 runs per strategy.

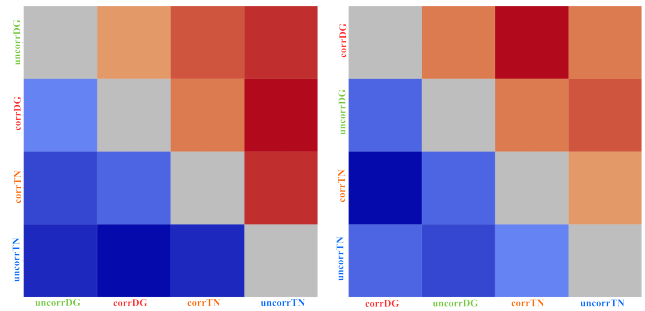

Figure 11: IESs applied to IQP: pairwise comparisons among the four IESs using the `IOHanalyzer` tool [31]. The heatmaps are generated according to the fraction of times the strategies' mean values are better over all the considered functions in 64D – with red reflecting total dominance versus blue reflecting total underperformance. The calculations are conducted in isolation for both the separable (LEFT) and non-separable (RIGHT) subsets. The strategies are ordered according to their obtained ranking, so the reader must give mind to the associated labels per each row/column. Importantly, the `uncorrDG` strategy dominates the separable subset with the `corrDG` as a first runner-up, whereas their ranks are flipped when shifting to the non-separable subset – consistently across dimensions (32D and 128D are not shown here).

pattern also held true for 32D and 128D, allowing the study to draw clear comparative performance conclusions on this testbed:

- The uncorrelatedDG IES dominates the separable subset; the `correlatedDG` IES is always the first runner-up.
- The `correlatedDG` IES dominates the nonseparable subset; the `uncorrelatedDG` IES is always the first runner-up.
- DG-based IESs always outperform TN-based IESs over the tested integer suite.

6 SUMMARY AND CONCLUSIONS

This study examined integer mutations for ESs when operating in unbounded search spaces, subject to either independence or correlations. We first clarified both the principal similarities and differences between integer and continuous unconstrained optimization from an evolutionary-algorithms perspective. Pointing out that even a convex quadratic model can exhibit multimodality once its domain is discretized to the integer lattice (see Supplementary Material), we motivated the need for mutation operators that respect the geometry of the discrete space. Central to this discussion is our proposed shift of attention to ℓ_1-norm symmetries: whereas classical analyses typically assume ℓ_2-invariance, we conceptually demonstrated (Figures 1 and 3) that the combinatorial setting can benefit from symmetry with respect to Manhattan distances. Building on this

insight, we proposed a procedure for drawing correlated integer samples that exactly follow the DG distribution, as proposed for independent sampling by Rudolph [22]. Visual inspections and correlation statistics confirm that the intended dependence structure is faithfully reproduced. At the same time, it is currently unclear how effectively the considered statistical measures can express an ℓ_1-norm-driven correlation in a meaningful sense – further investigation is required and will be explored in future research Moreover, a comparative study of candidate mutation kernels – including the TN and the DG distributions – highlighted the latter's favorable entropy profile; among all tested distributions it delivers also in the correlated case the largest entropy for a given expected step length, thus offering the richest exploratory power. We mentioned empirical tests on benchmark IQP models, recently reported in [25], which reveal significantly better convergence when the IES employs the DG-based mutation operator. Yet every configuration, irrespective of the mutation kernel, experiences an apparently unavoidable loss of accuracy once the search enters the immediate vicinity of the optimum. The resulting stagnation sets in almost deterministically at a characteristic distance, reminiscent of a *phase transition*, while simple tweaks fail to remove it.

Looking ahead, several avenues merit closer attention. A rigorous run-time analysis, especially of the final search phase, could uncover the mechanisms behind the observed stagnation and inspire strategies that maintain progress to exact optima. Because the DG law generates step lengths in closed form, it is a natural candidate for derandomized step-size adaptation schemes such as the CMA-ES; combining the DG distribution might even further improve results. Further experiments on problems with correlated integer decision variables – NK landscapes [15] provide an immediate MI testbed, as well as real-world instances – will demonstrate how well the method generalizes to MI cases. In the MI case it might however be difficult to capture the dependencies among integer and continuous variables [7]. Finally, enriching the algorithm with boundary-aware mutations should enable effective search when true optima lie on, or close to, integer constraints, and extending the framework to MI settings will widen its practical reach.

REFERENCES

[1] Thomas Bäck. 1996. *Evolutionary Algorithms in Theory and Practice.* Oxford University Press, New York, NY, USA.

[2] Thomas Bäck, Christophe Foussette, and Peter Krause. 2013. *Contemporary Evolution Strategies.* Springer-Verlag Berlin Heidelberg.

[3] Thomas Bäck and Martin Schütz. 1995. Evolution Strategies for Mixed Integer Optimization of Optical Multilayer Systems. In *Evolutionary Programming IV – Proc. Fourth Annual Conf. Evolutionary Programming*. The MIT Press, 33–51.

[4] Hans-Georg Beyer. 2001. *The Theory of Evolution Strategies.* Springer, Heidelberg.

[5] Jacob de Nobel, Diederick Vermetten, Hao Wang, Anna V. Kononova, Günter Rudolph, and Thomas Bäck. 2025. Abnormal Mutations: Evolution Strategies Don't Require Gaussianity. arXiv:2502.03148 [cs.NE]

[6] Michael Emmerich, Monika Grötzner, Bernd Groß, and Martin Schütz. 2000. Mixed-integer evolution strategy for chemical plant optimization with simulators. In *Evolutionary Design and Manufacture: Selected Papers from ACDM'00.* Springer, 55–67.

[7] Michael TM Emmerich, Rui Li, Anyi Zhang, Ildikó Flesch, and PJF Lucas. 2008. Mixed-integer bayesian optimization utilizing a-priori knowledge on parameter dependences. (2008).

[8] Ryoki Hamano, Shota Saito, Masahiro Nomura, and Shinichi Shirakawa. 2022. CMA-ES with margin: lower-bounding marginal probability for mixed-integer black-box optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference* (Boston, Massachusetts) *(GECCO '22)*. ACM, New York, NY, USA, 639–647. https://doi.org/10.1145/3512290.3528827

[9] Nikolaus Hansen. 2000. Invariance, Self-Adaptation and Correlated Mutations in Evolution Strategies. In *Parallel Problem Solving from Nature PPSN VI*, Marc Schoenauer, Kalyanmoy Deb, Günther Rudolph, Xin Yao, Evelyne Lutton, Juan Julian Merelo, and Hans-Paul Schwefel (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 355–364.

[10] Nikolaus Hansen, Anne Auger, Raymond Ros, Steffen Finck, and Petr Pošík. 2010. Comparing Results of 31 Algorithms from the Black-box Optimization Benchmarking BBOB-2009. In *Proceedings of the 12th Annual Conference Companion on Genetic and Evolutionary Computation* (Portland, Oregon, USA) *(GECCO '10)*. ACM, New York, NY, USA, 1689–1696. https://doi.org/10.1145/1830761.1830790

[11] Nikolaus Hansen, Fabian Gemperle, Anne Auger, and Petros Koumoutsakos. 2006. When do heavy-tail distributions help?. In *International Conference on Parallel Problem Solving from Nature*. Springer, 62–71.

[12] Yuan Hong and Dirk Arnold. 2023. Evolutionary Mixed-Integer Optimization with Explicit Constraints. In *Proceedings of the Genetic and Evolutionary Computation Conference* (Lisbon, Portugal) *(GECCO '23)*. ACM, New York, NY, USA, 822–830. https://doi.org/10.1145/3583131.3590467

[13] Ravi Kannan and László Lovász. 1988. Covering Minima and Lattice-Point-Free Convex Bodies. *Annals of Mathematics* 128, 3 (1988), 577–602. https://doi.org/10.2307/1971436

[14] R.C. Kelahan and J.L. Gaddy. 1978. Application of the Adaptive Random Search to Discrete and Mixed Integer Optimization. *Internat. J. Numer. Methods Engrg.* 12 (1978), 289–298.

[15] Rui Li, Michael Emmerich, Jeroen Eggermont, Thomas Bäck, Martin Schütz, Jouke Dijkstra, and Johan Reiber. 2013. Mixed integer evolution strategies for parameter optimization. *Evolutionary Computation* 21, 1 (2013), 29–64.

[16] Tristan Marty, Yann Semet, Anne Auger, Sébastien Héron, and Nikolaus Hansen. 2023. Benchmarking CMA-ES with Basic Integer Handling on a Mixed-Integer Test Problem Suite. In *Proceedings of the Companion Conference on Genetic and Evolutionary Computation* (Lisbon, Portugal) *(GECCO '23 Companion)*. ACM, New York, NY, USA, 1628–1635. https://doi.org/10.1145/3583133.3596411

[17] Christos H. Papadimitriou and Kenneth Steiglitz. 1998. *Combinatorial Optimization: Algorithms and Complexity.* Dover Publications, Mineola, NY, USA.

[18] Ingo Rechenberg. 1973. *Evolutionsstrategies: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution.* Frommann-Holzboog Verlag, Stuttgart, Germany.

[19] Edgar Reehuis, Johannes Kruisselbrink, André Deutz, Thomas Bäck, and Michael Emmerich. 2011. Multiobjective optimization of water distribution networks using SMS-EMOA. *Evolutionary Methods for Design, Optimisation and Control with Application to Industrial Problems (EUROGEN 2011)* (2011), 269–279.

[20] Victor Reis and Thomas Rothvoss. 2024. The Subspace Flatness Conjecture and Faster Integer Programming. arXiv:2303.14605 [math.OC] https://arxiv.org/abs/2303.14605

[21] Günther Rudolph. 1992. On Correlated Mutations in Evolution Strategies. In *Parallel Problem Solving from Nature - PPSN II.* Elsevier, Amsterdam, 105–114.

[22] Günter Rudolph. 1994. An Evolutionary Algorithm for Integer Programming. In *Parallel Problem Solving from Nature-PPSN III.* Springer, 139–148.

[23] Alexander Schrijver. 1998. *Theory of Linear and Integer Programming.* John Wiley and Sons, Chichester, England.

[24] Hans-Paul Schwefel. 1995. *Evolution and Optimum Seeking.* John Wiley & Sons, Inc., New York, NY, USA.

[25] Ofer M. Shir and Michael Emmerich. 2025. Correlated Geometric Mutations for Integer Evolution Strategies. In *Proceedings of the Companion Conference on Genetic and Evolutionary Computation (GECCO '25 Companion)*. ACM, New York, NY, USA. https://doi.org/10.1145/3712255.3734283

[26] Ofer M. Shir and Michael Emmerich. 2025. Multiobjective Mixed-Integer Quadratic Models: A Study on Mathematical Programming and Evolutionary Computation. *IEEE Transactions on Evolutionary Computation* 29, 3 (2025), 661–675. https://doi.org/10.1109/TEVC.2024.3374519

[27] Ofer M. Shir, Boris Yazmir, Assaf Israeli, and Dan Gamrasni. 2022. Algorithmically-Guided Postharvest Protocols by Experimental Combinatorial Optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion* (Boston, Massachusetts) *(GECCO '22)*. ACM, New York, NY, USA, 2027–2035. https://doi.org/10.1145/3520304.3533976

[28] Ofer M. Shir and Amir Yehudayoff. 2020. On the Covariance-Hessian Relation in Evolution Strategies. *Theoretical Computer Science* 801 (2020), 157–174. https://doi.org/10.1016/j.tcs.2019.09.002

[29] Robert Tibshirani. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58, 1 (1996), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[30] Tea Tušar, Dimo Brockhoff, and Nikolaus Hansen. 2019. Mixed-integer benchmark problems for single- and bi-objective optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference* (Prague, Czech Republic) *(GECCO '19)*. ACM, New York, NY, USA, 718–726. https://doi.org/10.1145/3321707.3321868

[31] Hao Wang, Diederick Vermetten, Furong Ye, Carola Doerr, and Thomas Bäck. 2022. IOHanalyzer: Detailed Performance Analyses for Iterative Optimization Heuristics. *ACM Trans. Evol. Learn. Optim.* 2, 1, Article 3 (April 2022). https://doi.org/10.1145/3510426