Statistics of Partial Minima

E. Ben-Naim[1],* M. B. Hastings[1],† and D. Izraelevitz[2]‡
[1] *Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545 USA*
[2] *Decision Applications Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545 USA*

We study pseudo-optimal solutions to multi-objective optimization problems by introducing partial minima defined as follows. Point $\mathbf{x}$ k-dominates $\mathbf{x}'$ when at least k of the coordinates of $\mathbf{x}$ are smaller than the corresponding coordinates of $\mathbf{x}'$. A point not k-dominated by any other point in the set is a k-minimum or a partial minimum, generalizing the global minimum. We study statistical properties of partial minima for a set of N points independently distributed inside the d-dimensional unit hypercube using exact probabilistic methods and heuristic scaling techniques. The average number of partial minima, A, decays algebraically with the total number of points, $A \sim N^{-(d-k)/k}$, when $1 \leq k < d$. Interestingly, there are $k-1$ distinct scaling laws characterizing the largest coordinates: the distribution $P(y_j)$ of the jth largest coordinate, y_j, decays algebraically, $P(y_j) \sim (y_j)^{-\alpha_j - 1}$, with $\alpha_j = j\frac{d-k}{k-j}$ for $1 \leq j \leq k-1$. The average number of partial minima grows logarithmically, $A \simeq \frac{1}{(d-1)!}(\ln N)^{d-1}$, when $k = d$. The full distribution of the number of minima is obtained in closed form in two-dimensions.

PACS numbers: 02.50.Cw, 05.40.-a, 89.20.Ff, 89.75.Da

A host of decisions in computer science, economics, politics, and everyday life involve multiple criteria or multiple objectives [1–4]. A pedestrian choosing a walking path considers the distance, the number of turns, and the number of traffic lights. In business, takeover bids are decided on a multitude of complex conditions in addition to the total monetary offer. In elections, voters examine how candidates stand on multiple issues.

In multi-objective optimization, a solution that is optimal with respect to all criteria is rarely possible and instead, one faces a set of choices that are suboptimal on at least one criterion. Decisions require algorithms to weed out clearly inferior choices, sort through all the remaining imperfect choices, and evaluate the relative trade-offs between costs.

Since in multi-objective optimization, a global optimum is unlikely, we are interested in identifying points that are close to optimum. In this paper, we propose a pseudo-optimality criterion, and derive the likelihood of finding pseudo-optimal solutions as a function of the number of choices.

By definition, a global optimum is superior in all cost dimensions. Intuitively, one may define a pseudo-optimum as superior to all alternatives along a large number of cost dimensions. For example in a three-cost scenario, there may not be any choice that is optimal with respect to all three costs, but we may be able to find choices that are better than any alternative along two costs, see Figure 1. In a voting scenario, no candidate may have the most attractive position on all issues to a given voter. In this case, a voter might naturally restrict her attention to a candidate, or candidates, who have the most attractive position on as many issues as possible.

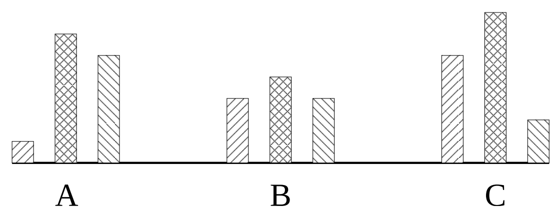

FIG. 1: Illustration of partial domination and partial minima. There are three costs associated with A, B, and C. A and C dominate B on one cost and thus, B is not a 1-minimum. Neither A nor C dominate B on two costs and thus, B is a 2-minimum.

Let us represent our choices as N points in d-dimensions with coordinates $\mathbf{x} \equiv (x_1, x_2, \ldots, x_d)$. Each coordinate $x_i > 0$ represents a distinct cost. By convention, small-x values are superior and are considered *dominant*. Partial minima, a formalization of the pseudo-optima concept discussed above, are defined as follows. A point $\mathbf{x}$ is said to be *k-dominated* by $\mathbf{x}'$ when at least k of the coordinates of $\mathbf{x}$ are larger then the corresponding coordinates of $\mathbf{x}'$. A point is said to be a partial minimum, or formally a *k-minimum*, when it is not k-dominated by any other point in the set as illustrated in Figure 1. We stress that a partial minimum is not required to dominate all other points on the same $d-k$ coordinates and may dominate different points along different coordinates. The parameter $1 \leq k \leq d$ quantifies the quality of the partial minimum: a smaller k value represents a more stringent condition. The two extremes are the global minimum, $k = 1$, where every coordinate is a minimum of the point set, and the efficient set, $k = d$, that includes all points that are not obviously dominated by other points as shown later in Figure 3. Partial min-

*Electronic address: ebn@lanl.gov
†Electronic address: hastings@lanl.gov
‡Electronic address: izraelevitz@lanl.gov

ima are conditional multivariate extrema and their properties are amenable to analysis using a statistical physics perspective [5–9].

In this study, we obtain exact statistical properties of partial minima including the multivariate density and its asymptotic behavior as well as scaling properties such as the typical size and average number of minima. We present two major results. First, as a function of the set size N, the average number of minima decays algebraically when $1 \leq k < d$, and grows logarithmically when $k = d$. Second, there are $k - 1$ different scaling laws for the largest coordinates, each following a power-law distribution with $k-1$ distinct exponents. The rest of the $d+1-k$ coordinates are characterized by distributions with sharp tails. We also discuss the relevance of these results to the multi-objective shortest path on graphs, a central problem in multi-objective optimization.

We consider the situation where there are no correlations between the coordinates. That is, each coordinate is independently drawn from some distribution. As discussed below, this situation is equivalent to a uniform distribution in the unit hypercube. Thus, we conveniently assume that x_i is uniformly distributed in $[0 : 1]$ for all $1 \leq i \leq d$.

Heuristic Arguments. Elementary scaling laws for the typical size of a partial minimum and the average number of minima are derived heuristically. We assume that (i) the partial minimum is dominant on a fixed set of k coordinates, and (ii) all its coordinates are equal, $x_i = x$, for all i. By the partial minimum definition, the corresponding k-dimensional hypercube contains only the partial minimum itself. The volume of this hypercube is x^k and the expected number of points inside this hypercube must be of order one, $Nx^k \sim 1$. Consequently, the typical size x decays algebraically with N,

$$x \sim N^{-\frac{1}{k}}. \tag{1}$$

This characteristic scale decreases as the minimum condition becomes more stringent, that is, as k decreases.

The expected number of partial minima

$$A \sim N^{-\frac{d-k}{k}} \tag{2}$$

follows from the expected number of points inside the d-dimensional hypercube with linear dimension x, Nx^d. Partial minima are asymptotically rare and the scale (1) decays indefinitely. Furthermore, with a small probability, there is only one minimum when N is large. The scaling estimate (2) coincides with the exact value $A = N^{-(d-1)}$ for $k = 1$, since any point is a global minimum with probability N^{-d}. For $k = d$, the minimum in any one coordinate is a partial minimum and thus, there is at least one partial minimum. Indeed, the decay exponent $\frac{d-k}{k}$ in (2) vanishes. This special case is discussed separately.

The Density of Minima. The density $P_{d,k}(\mathbf{x})$ of k-minima located at $\mathbf{x}$ is obtained analytically through a formal generalization of the heuristic argument above. For example, in two dimensions the density is

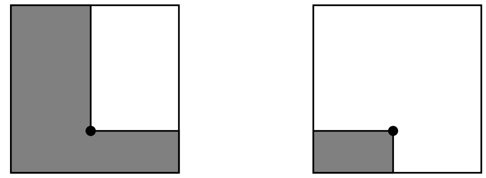

FIG. 2: Illustration of the excluded area for a global minimum ($k = 1$, left) and points on the efficient set ($k = 2$, right) in two-dimensions. Points in the gray region k-dominate the distinguished point.

$$P_{2,k}(x_1, x_2) = \begin{cases} N\left[1 - (x_1 + x_2 - x_1x_2)\right]^{N-1} & k = 1, \\ N\left[1 - x_1x_2\right]^{N-1} & k = 2. \end{cases}$$

The factor N is the number of ways to choose the minimum, and the second factor guarantees that the rest of the points do not dominate the minimum at (x_1, x_2). These points must not fall inside an L-shaped region of area $x_1 + x_2 - x_1x_2$ or equivalently $1 - (1 - x_1)(1 - x_2)$ when $k = 1$ or a rectangle of area x_1x_2 when $k = 2$ as illustrated in Figure 2.

In general, the density of minima

$$P_{d,k}(\mathbf{x}) = N[1 - G_{d,k}(\mathbf{x})]^{N-1}. \tag{3}$$

reflects that the $N - 1$ points are excluded from a d-dimensional region of volume $G_{d,k}(\mathbf{x})$. The excluded volume obeys the recursion

$$G_{d,k}(\mathbf{x}) = x_d\, G_{d-1,k-1}(\mathbf{x}) + (1 - x_d)\, G_{d-1,k}(\mathbf{x}). \tag{4}$$

In our notation, the dimensional index of a function dictates the dimension of its vectorial argument so the vectors on the right hand side of (4) have $d-1$ components. We obtain the recursion relation (4) by separating the excluded region into two regions: one in which the dth coordinate is dominant and one in which it is not. Using the boundary conditions $G_{d,0} = 1$ and $G_{d,k} = 0$ when $k > d$, we recover $G_{1,1} = x_1$ and $G_{2,1} = x_1 + x_2 - x_1x_2$. Furthermore,

$$G_{3,k} = \begin{cases} 1 - (1 - x_1)(1 - x_2)(1 - x_3) & k = 1, \\ x_1x_2 + x_1x_3 + x_2x_3 - 2x_1x_2x_3 & k = 2, \\ x_1x_2x_3 & k = 3. \end{cases}$$

In general, $G_{d,d} = \prod_{i=1}^{d} x_i$ and $G_{d,1} = 1 - \prod_{i=1}^{d}(1 - x_i)$.

Scaling. In the limit $N \to \infty$, the product term x_1x_2 in $P_{2,1} = N[1 - (x_1 + x_2 - x_1x_2)]^{N-1}$ is negligible compared with the linear term $x_1 + x_2$ and thus,

$$P_{2,1}(x_1, x_2) \to Ne^{-N(x_1+x_2)}.$$

Generally, only the kth degree terms are asymptotically relevant and the leading behavior is

$$P_{d,k}(\mathbf{x}) \to Ne^{-NF_{d,k}(\mathbf{x})}. \tag{5}$$

The auxiliary function $F_{d,k}(\mathbf{x})$ contains $\binom{d}{k}$ terms, each a distinct product of degree k. For example,

$$F_{3,k} = \begin{cases} x_1 + x_2 + x_3 & k = 1, \\ x_1x_2 + x_1x_3 + x_2x_3 & k = 2, \\ x_1x_2x_3 & k = 3. \end{cases}$$

The auxiliary function equals the sum, $F_{d,1} = \sum_{i=1}^{d} x_i$, and the product, $F_{d,d} = \prod_{i=1}^{d} x_i$, in the two extremes. The function $F_{d,k}(\mathbf{x})$ is defined recursively

$$F_{d,k}(\mathbf{x}) = x_d F_{d-1,k-1}(\mathbf{x}) + F_{d-1,k}(\mathbf{x}) \tag{6}$$

for $1 \leq k \leq d$ with the boundary condition $F_{0,k} = \delta_{k,0}$. This recursion follows from (4) by dropping the higher-degree term $x_d\, G_{d-1,k}(\mathbf{x})$.

The asymptotic behavior (5) can be recast in the scaling form

$$P_{d,k}(\mathbf{x}) \to N\Phi_{d,k}(\mathbf{z}), \tag{7}$$

as $N \to \infty$. The scaling variable is $\mathbf{z} = \mathbf{x}N^{1/k}$, in accord with (1), and the scaling function is

$$\Phi_{d,k}(\mathbf{z}) = e^{-F_{d,k}(\mathbf{z})}. \tag{8}$$

The average number of k-minima equals the integral of the density, $A_{d,k} = \int d\mathbf{x}\, P_{d,k}(\mathbf{x})$, where $\int d\mathbf{x} \equiv \prod_{i=1}^{d} \int_0^1 dx_i$ [10]. When $k < d$, the asymptotic behavior of the average follows from the scaling form (7), $A_{d,k} \simeq a_{d,k}\, N^{-\frac{d-k}{k}}$, and is in agreement with (2). The proportionality constant $a_{d,k}$ equals the integral of the scaling function, $a_{d,k} = \int d\mathbf{z}\, \Phi_{d,k}(\mathbf{z})$, although now, the integration range is unrestricted, $\int d\mathbf{z} \equiv \prod_{i=1}^{d} \int_0^\infty dz_i$. The prefactor is trivial for perfect minima, $a_{d,1} = 1$, and otherwise, it can be obtained analytically only in a few exceptional cases.

Extreme Statistics. Since global minima are constrained along all cost coordinates, extremely large costs are exponentially rare, whenever such a global minimum exists. Because we have relaxed the minimality condition, this may not necessarily be the case for partial minima. In our voting example, a candidate who is attractive to a voter on a multitude of issues may be extremely unattractive on a particular one. How likely is such a scenario?

We begin our study of extremal statistics [11–13] by first considering the distribution of the largest coordinate in a partial minimum. Without loss of generality, we order the coordinates $x_1 < x_2 < \cdots < x_{d-1} < x_d$. Our focus is on the tail of the distribution of the variable x_d, corresponding to the regime $x_d \gg x_{d-1}$. We also restrict our attention to the limit $N \to \infty$. The distribution $Q_1(x_d)$ of the largest coordinate x_d equals the integral of the multivariate distribution with respect to the rest of the coordinates,

$$\begin{aligned} Q_1(x_d) &= \int dx_1 \cdots \int dx_{d-1}\, P_{d,k}(x_1, x_2, \cdots, x_d) \\ &\sim \int dx_1 \cdots \int dx_{d-1}\, N e^{-NF_{d,k}(\mathbf{x})} \\ &\sim \int dx_1 \cdots \int dx_{d-1}\, N e^{-Nx_d F_{d-1,k-1}(\mathbf{x})} \\ &\sim N^{-\frac{d-k}{k-1}} (x_d)^{-\frac{d-k}{k-1}-1}. \end{aligned} \tag{9}$$

The second line is obtained by substituting the leading asymptotic behavior (5) and the third line reflects that only the first term in (6) is relevant when $x_d \gg x_i$ for all $i < d$. Our last step is to multiply and divide the third line by x_d and then invoke the scaling law (2) for the average number of $k-1$-minima in $d-1$ dimensions. In essence, we utilize the fact that when one of the coordinates is very large, the partial minima criterion involves one less constraint in one less dimension [14]. The power-law decay of the distribution (9) shows that there is a substantial likelihood that x_d is relatively large.

The distribution $Q_2(x_{d-1})$ of the second largest coordinate x_{d-1} is obtained using the bivariate distribution $\tilde{Q}(x_{d-1}, x_d)$,

$$\begin{aligned} \tilde{Q}(x_{d-1}, x_d) &= \int dx_1 \cdots \int dx_{d-2}\, P_{d,k}(x_1, x_2, \cdots, x_d) \\ &\sim \int dx_1 \cdots \int dx_{d-2}\, N e^{-NF_{d,k}(\mathbf{x})} \\ &\sim \int dx_1 \cdots \int dx_{d-2}\, N e^{-Nx_{d-1}x_d F_{d-1,k-1}(\mathbf{x})} \\ &\sim N^{-\frac{d-k}{k-2}} (x_{d-1}x_d)^{-\frac{d-k}{k-2}-1}. \end{aligned} \tag{10}$$

The distribution $Q_2(x_{d-1})$ equals the integral of the bivariate distribution with respect to the largest coordinate, $Q_2(x_{d-1}) = \int_{x_{d-1}}^{1} dx_d\, \tilde{Q}(x_{d-1}, x_d)$. This integral is dominated by the divergence at the lower limit of integration, and consequently

$$Q_2(x_{d-1}) \sim N^{-\frac{d-k}{k-2}} (x_{d-1})^{-2\frac{d-k}{k-2}-1}. \tag{11}$$

The power-law tail is now steeper.

A similar calculation applies to the distributions of the $k-1$ largest elements. In general, the distribution $Q_j(y_j)$ of the jth largest element, y_j, with the definition $y_j \equiv x_{d+1-j}$, decays as a power-law,

$$Q_j(y_j) \sim N^{-\frac{d-k}{k-j}} (y_j)^{-\alpha_j - 1} \tag{12}$$

for $1 \leq j \leq k-1$. The decay exponent increases monotonically with the index j,

$$\alpha_j = j\,\frac{d-k}{k-j}. \tag{13}$$

We can verify the decay law (2) using $A \sim \int_{N^{-1/k}}^{1} dy_j\, Q_j(y_j)$ where the lower limit of integration is set by the typical size scale (1). Interestingly, there are $k-1$ distinct scaling behaviors for the $k-1$ largest elements. Each of these extremal coordinates is distributed according to a power-law distribution that is characterized by a distinct exponent.

This multiscaling behavior affects the behavior of the moments $\langle y_j^m\rangle$ defined as follows, $\langle y_j^m\rangle = I_m/I_0$, where $I_m = \int_{N^{-1/k}}^1 dy_j\, y_j^m\, Q_j(y_j)$. The integral I_m is dominated by the divergence at the lower cutoff when the order is small, $m\leq\alpha_j$, but otherwise, the integral I_m is finite. Consequently, the moments have the following scaling dependences on N

$$\langle y_j^m\rangle \sim \begin{cases} N^{-m/k} & m<\alpha_j,\\ \frac{1}{k}N^{-\frac{j(d-k)}{k(k-j)}}\ln N & m=\alpha_j,\\ N^{-\frac{j(d-k)}{k(k-j)}} & m>\alpha_j.\end{cases} \tag{14}$$

Low order moments exhibit ordinary scaling behavior as they are characterized by the typical size scale (1) that underlies the multivariate distribution function (8). As usual, there is a logarithmic correction at the crossover. High-order moments plateau at a fixed value that is independent of the index m, an indication that there is a significant probability that the extreme elements are of order one. Interestingly, the average size of the different coordinates may follow different scaling laws. For example, there are two scaling laws, $\langle y_1\rangle\sim N^{-1/6}$ and $\langle y_2\rangle\sim N^{-1/3}$ when $d=4$ and $k=3$. Of course, the sum $\sum_{i=1}^d x_i$ has the same extremal statistics as does x_d.

The crossover moment or equivalently the exponent α_j diverges as $k\to j$. Therefore, the smallest $d+1-k$ coordinates exhibit the ordinary scaling behavior

$$\langle y_j^m\rangle \sim N^{-m/k} \tag{15}$$

for $k\leq j\leq d$ and all moments of the respective distribution functions must be finite. In these cases, the distribution functions Q_j have tails that are as sharp as or sharper than an exponential. In the aforementioned case $d=4$ and $k=3$, the third and the fourth largest coordinates exhibit the ordinary scaling, $\langle y_3\rangle\sim\langle y_4\rangle\sim N^{-1/3}$.

Efficient Sets. The set of points that are not dominated on all coordinates by any other point are partial minima when $k=d$ (Figure 3). We refer to this set as the "efficient set". The efficient set, also termed the efficient frontier or Pareto equilibria, plays a central role in multi-objective optimization and has been studied extensively in economics, computer science, operations research, and game theory [15, 16]. Since there is no objective trade-off between costs, every point in the efficient set is potentially a solution to the multi-objective optimization problem. The study of the properties of efficient sets was the original motivation for our research.

In the special case $k=d$, the expected size of the efficient set, $E_d(N)\equiv A_{d,d}(N)$, obeys the recursion

$$E_d(N) = E_d(N-1) + \frac{1}{N}\,E_{d-1}(N). \tag{16}$$

The point with the largest x_d coordinate certainly does not dominate any other point. Furthermore, this point

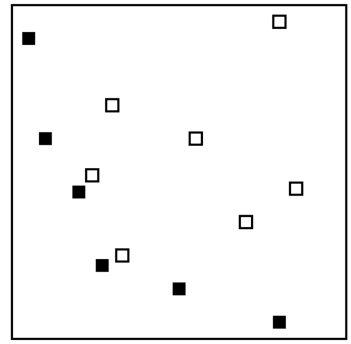

FIG. 3: Illustration of the efficient set in two-dimensions. Filled squares are on the efficient set and unfilled squares are not. Only four of the filled squares are on the convex hull.

is on the efficient set if and only if the rest of its $d-1$ coordinates are not dominated by any other point. This event occurs with probability $\frac{1}{N}E_{d-1}(N)$ and hence, the second term in the recursion. We note that the recursion (16) can also be obtained by performing the integration over x_d in $E_d(N) = N\int d\mathbf{x}\,[1-x_1x_2\cdots x_d]^{N-1}$. This integration is analytically feasible only if $k=1$ or $k=d$.

The recursion relation (16) is subject to the boundary condition $E_1(N)=1$. In two dimensions,

$$E_2(N) = 1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{N}, \tag{17}$$

or alternatively, $E_2(N)=H(N)$, where $H(N)=\sum_{n=1}^N\frac{1}{n}$ is the harmonic number. The average size of the efficient set grows logarithmically, $E_2(N)=\ln N+\gamma+\cdots$ where $\gamma=0.57721$ is Euler's constant. In three dimensions, we have $E_3(N)=\sum_{n=1}^N\frac{1}{n}H(n)$, and asymptotically, $E_3(N)\simeq\frac{1}{2}(\ln N)^2$. The large-$N$ behavior is obtained in general by converting the difference equation (16) into a differential equation $dE_d/dN = E_{d-1}/N$. The expected size of the efficient set grows logarithmically,

$$E_d(N)\simeq\frac{1}{(d-1)!}(\ln N)^{d-1}. \tag{18}$$

This logarithmic growth reflects that the integral of the scaling function, $\int d\mathbf{z}\,\Phi_{d,d}(\mathbf{z})$, is divergent at the upper limit. A straightforward generalization of the calculation above shows that the distribution of the extremal coordinates has a logarithmic correction,

$$Q_j(y_j)\sim(\ln N)^{d+1-j}(y_j)^{-1}|\ln y_j|^{j-1}, \tag{19}$$

for $1\leq j\leq d-1$. We can verify that the average number of points is consistent with the exact behavior $\int_{N^{-1/d}} dy\,Q_j(y_j)\sim(\ln N)^{d-1}$ as in (18). The crossover moment vanishes and the moments decay logarithmically,

$$\langle y_j^m\rangle\sim(\ln N)^{-j}, \tag{20}$$

where $m>0$ and $1\leq j\leq d-1$.

Two-dimensions. For the case $d=2$, we obtain closed form expressions for the distribution function of partial minima. This permits us to establish central limit-type behaviors for the distribution of the size of the efficient set.

In two-dimensions, the distribution function $p_n(N)$ for the event that the efficient set includes n points, where $1 \le n \le N$, satisfies the recursion [17]

$$p_n(N) = \left(1 - N^{-1}\right) p_n(N-1) + N^{-1} p_{n-1}(N-1) \quad (21)$$

and is subject to the boundary condition $P_n(0) = \delta_{n,0}$. On the square, there are two coordinates: x_1 and x_2. We can derive (21) by alluding to the same reasoning behind (16), i.e., the point with the largest x_2 coordinate will be on the efficient set if and only if its x_1 coordinate is minimal, an event that occurs with probability N^{-1}.

Recursion equations for the average $E(N) = \langle n \rangle$ and the variance $V(N) = \langle n^2 \rangle - \langle n \rangle^2$ with $\langle f(n) \rangle \equiv \sum_{n=1}^{N} f(n) P_n$ are obtained by summing (21). The average satisfies $E(N) = E(N-1) + N^{-1}$ in accord with (16) and the variance satisfies $V(N) = V(N-1) + N^{-1} - N^{-2}$. Thus, the variance equals the difference between the first and the second harmonic numbers

$$V(N) = H(N) - H^{(2)}(N) \quad (22)$$

where $H^{(2)}(N) = \sum_{n=1}^{N} n^{-2}$. The variance and the average have identical leading asymptotic behaviors, $V(N) = \ln N + (\gamma - \frac{1}{6}\pi^2) + \cdots$.

With the transformation $p_n(N) = \frac{1}{N!}\tilde{p}_n(N)$, the auxiliary function $\tilde{p}_n(N)$ satisfies the recursion

$$\tilde{p}_n(N) = (N-1)\tilde{p}_n(N-1) + \tilde{p}_n(N-1) \quad (23)$$

with $\tilde{p}_n(0) = \delta_{n,0}$. This recursion defines the Stirling numbers $\left[{N \atop n}\right]$ [18] so $\tilde{p}_n(N) = \left[{N \atop n}\right]$. Therefore, the full probability distribution is expressed in closed form,

$$p_n(N) = \frac{1}{N!}\begin{bmatrix} N \\ n \end{bmatrix} \quad (24)$$

for $0 \le n \le N$.

The general asymptotic behavior, derived in [19],

$$p_n(N) \simeq \frac{1}{N}\frac{1}{\Gamma(n/\ln N)}\frac{(\ln N)^n}{n!} \quad (25)$$

applies in the limit $n \to \infty$ $N \to \infty$ with the ratio $n/\ln N$ finite. For small $n \ll \ln N$, the distribution is Poissonian, $P_n(N) = N^{-1}(\ln N)^{n-1}/(n-1)!$ and for large n, the distribution approaches a Gaussian centered at the average $E(N) \simeq \ln N$ with the variance $V(N) \simeq \ln N$,

$$p_n(N) \to \frac{1}{\sqrt{2\pi \ln N}} \exp\left[-\frac{(n-\ln N)^2}{2\ln N}\right]. \quad (26)$$

We note that the convex hull, a subset of the efficient set (see Figure 3), is characterized by similar statistical properties including a limiting Gaussian distribution and logarithmic growths, albeit with different prefactors, of the average and the variance [20–22].

Multi-Objective Shortest Path. The multi-objective shortest path on a graph is defined as follows. Consider a graph, possibly with multiple edges connecting pairs of nodes, with d different costs on each edge. Fix the source and the destination nodes, and then consider all paths from source to destination, assigning d total costs to each path computed as the sum of the d individual costs of the path's constituent edges. The multi-objective shortest path problem consists of finding the efficient set of paths. Generally, finding the efficient set is an NP-hard problem, although less demanding approximation schemes exist [23, 24]. Nevertheless, the computation time of the approximation scheme depends crucially on the size of the efficient set.

Suppose the edge costs are independent, random draws from a common distribution. We can consider two limiting topologies. First, for a graph of two nodes connected by N edges, the number of elements in the efficient set grows poly-logarithmically in the number of edges as shown in (18). Second, for a one-dimensional chain of nodes where each pair of neighboring nodes is connected by a pair of edges, the total path costs become correlated [23], even though the individual edge costs are not. We have conducted numerical studies that find that the size of the efficient set is highly sensitive to the distribution of edge costs. Assuming each edge has two costs, (w_1, w_2), both chosen from some continuous distribution, the convex hull grows linearly in the length of the chain. Interestingly, we observed various behaviors for the size of the efficient set, ranging from linear in the length of the chain, to power law behavior, characterized by exponents greater than unity, up to stretched exponential behavior.

Finally, we consider Erdös-Renyi random graphs [25–27]. Using the fact that the shortest path between two randomly chosen nodes grows logarithmically with the total number of nodes in the graph and the fact that paths that are close in length to the shortest path weakly overlap and hence their costs are weakly correlated, the results in this paper can be used to heuristically show [28] that the size of the efficient set of paths grows poly-logarithmically with the number of nodes as in (18). This number is much smaller than the number found for chains where the paths are correlated.

Conclusions. We proposed partial minima as a protocol for identifying pseudo optimal solutions to multi-objective optimization problems. Partial minima are defined by a parameter k: a point in d-dimensions that dominates all other points on at least $d-k$ coordinates is a partial minimum. As this optimality criterion becomes more stringent, partial minima improve in quality but are less probable. In the extreme case $k = d$, the number of partial minima grows logarithmically with the total number of points.

Remarkably, there is a series of distinct power-law distributions that characterize the largest coordinates with a consequent multiscaling distribution of the moments, while the rest of the coordinates obey ordinary scaling. Viewed as quasi-optimal solutions to multi-objective optimization problems, partial minima involve a trade-off. When the optimality criterion is relaxed, these quasi-optima become more likely, but are more likely to incur at least one extremely large cost.

Our results hold as long as the set of points are not correlated, that is, as long as they are drawn from independent distributions. These distributions need not be identical. If the ith coordinate is drawn from the distribution $f_i(x_i)$, the transformation $x_i \to \int_0^{x_i} dy_i f_i(y_i)$ and $dx_i \to f_i(x_i)dx_i$, maps to a uniform distribution in the unit hypercube. Correlations present an interesting challenge and we anticipate serious modifications to the scaling laws above. For instance, it is simple to show that the size of the efficient set grows as a power of the number of points, $\sim N^{1/2}$, rather than a logarithm, when the points are uniformly distributed inside the unit circle. Incidentally, this growth is much faster than the $N^{1/3}$ for the corresponding number of points in the convex hull [20].

Another interesting issue is the crossover from the algebraic decay (2) to the logarithmic growth (18). The average number of partial minima decreases monotonically with N when k is small, but is a non-monotonic function of N when k is large. For example, when $d = 4$ and $k = 3$, the average $A_{d,k}$ peaks at $N = 16$. It will be interesting to elucidate how the height and the location of this peak scales with N.

Acknowledgments. We thank Gunes Ercal-Ozkaya for suggesting random graphs and Paul Krapivsky for useful discussions on convex hulls. We acknowledge financial support from DOE grant DE-AC52-06NA25396.

---

[1] R. E. Steuer, *Multiple Criteria Optimization: Theory, Computations, and Application* (Wiley, New York, 1986).
[2] Y. Sawaragi, H. Nakayama, and T. Tanino, *Theory of Multiobjective Optimization*, (Academic Press, Orlando, 1985).
[3] D. Fudenberg and J. Tirole, *Game Theory*, (MIT Press, Cambridge, 1983)
[4] M. J. Osborne and A. Rubenstein, *A Course in Game Theory*, (MIT Press, Cambridge, 1994).
[5] M. Mézard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
[6] B. Derrida and J. L. Lebowitz, Phys. Rev. Lett. **80**, 209 (1998).
[7] E. Ben-Naim, P. L. Krapivsky, and S. N. Majumdar, Phys. Rev. E **64**, 035101 (2001).
[8] G. Korniss, Z. Toroczkai, M. A. Novotny, and P. A. Rikvold, Phys. Rev. Lett. **84**, 1351 (2000).
[9] P. L. Krapivsky and S. Redner, Phys. Rev. Lett. **89**, 258703 (2002).
[10] For small N, these integrals can be calculated manually or through recursion formulas. For example, $A_{d,k} = \frac{1}{2}(A_{d-1,k-1} + A_{d-1,k-1})$ when $N = 2$.
[11] E. J. Gumbel, *Statistics of Extremes* (Columbia University Press, New York, 1958).
[12] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics* (R.E. Krieger Publishing Co., Malabar, 1987).
[13] R. E. Ellis, *Entropy, Large Deviations, and Statistical Mechanics* (Springer-Verlag, New York, 1985).
[14] Formally, when $x_d = 1$ then $G_{d,k}(\mathbf{x}) = G_{d-1,k-1}(\mathbf{x})$.
[15] T. Ottmann, E. Soisalon-Soininen, and D. Wood, Info. Sci. **33**, 157 (1984).
[16] E. Fink and D. Wood, Jour. Geometry **62**, 99 (1998).
[17] Generally, $p_1(N) = A_{d,1} = N^{-(d-1)}$.
[18] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science* (Reading, Mass.: Addison-Wesley, 1989).
[19] E. Ben-Naim and P. L. Krapivsky, J. Phys. A **37**, 5949 (2004).
[20] A. Rényi and R. Sulanke, Z. Wahrsch. Verw. Geb. **2**, 75–84 (1963).
[21] B. Efron, Biometrika, **52**, 331 (1965).
[22] V. Vu, Adv. Math. **207**, 221 (2006).
[23] A. Warburton, Oper. Res. **35**, 70 (1987).
[24] G. Tsaggouris and C. Zaroliagis, Lect. Notes in Comp. Sci. **4288**, 389 (2006).
[25] B. Bollobás, *Random Graphs* (Academic Press, London, 1985).
[26] S. Janson, T. Łuczak, and A. Rucinski, *Random Graphs* (John Wiley & Sons, New York, 2000).
[27] S. Janson, Combin. Probab. Comput. **8**, 347 (1999).
[28] E. Ben-Naim, M. B. Hastings, D. Izraelevitz, unpublished.