Optimization of Additional Information Acquisition in Decision Making Problems: Solution Methods

David Grace
Lehigh University

Eugene Perevalov
Lehigh University

Report: 12T-005

Optimization of Additional Information Acquisition in Decision Making Problems: Solution Methods

E. Perevalov and D. Grace

Abstract

When additional information sources are available in decision making problems that allow stochastic optimization formulations, an important question is how to optimally use the information the sources are capable of providing. A framework that relates information characteristics of a source to solution quality characteristics of the problem was proposed in a companion paper where the problem of optimal information acquisition was formulated as that of minimization of the expected loss of the solution subject to capacity constraints of the information source. In this paper, approximate solution methods for that problem are developed making use of probability metrics method and its application for scenario reduction in stochastic optimization.

1 Introduction

In many practically important decision making problems where uncertainty about input data is present and optimization methods are appropriate (due to a large number of possible solutions) sources of additional information are in principle available. Often, information that such sources possess fails to be taken advantage of due to its perceived and factual imprecision and to the lack of methodology that allows to do this in a controlled and regular fashion.

The main purpose of such methodology, as seen by us, is to be able to (i) estimate the amount of decision quality improvement that's possible for the given problem and with the help of a given amount – suitably defined – of additional information, (ii) given a specific information source, find the particular way the maximum possible improvement in decision quality can be achieved. Correspondingly, it appears logical – if such a methodology is to be developed – to begin with a quantitative description of information exchange between the analyst/decision maker and information source(s), then try to relate the information related measures to decision quality and, finally, to optimize the latter given suitable constraints (that describe limitations of the information source) on the former. The first (informational) component of such a methodology is addressed in [31, 30, 32]. The second (relating informational characteristics to decision quality) was explored in [33] where the main such relations were discussed and the problem of optimizing the solution quality given information source characteristics was formulated. The next logical step is to look for specific solution methods for this problem. This is the main topic of the present article.

The overall approach developed in this paper and papers cited above builds – directly and indirectly – on several bodies of previous research. The area of statistical decision making has dealt with the idea of improving solution quality by means of acquiring additional information. There have been applications to innovation adoption [28], [20], fashion decisions [11] and vaccine composition decisions for flu immunization [24] can be mentioned in this regard. Some authors

[10], [9] even introduced models (e.g. effective information model) for accounting for the actual, or effective, amount of information contained in the received observations. One could also mention the recent work on optimal decision making in the absence of the knowledge of the distribution shape and parameters [17, 25, 1]. The difference of the proposed approach is in that it explicitly describes and allows to optimize over not just the quantity of additional information but also its content and is based on explicit description of properties of information sources.

On the other hand, the proposed approach can be looked upon as an attempt to make Information Theory methods useful for optimization and decision making under uncertainty. The field of Information Theory, born from Shannon's work on the theory of communications [41] since had great success in a number of fields that include, besides communication theory, statistical physics [18, 19], computer vision [43], climatology [29, 42], physiology [21] and neurophysiology [3]. Generalized Information Theory (see e.g. [22], [27, 14]) addresses problems of characterizing uncertainty in frameworks that are more general than classical probability such as Dempster-Shafer theory [40]. A bit more specifically speaking, the approach developed here is based on a theory of information exchange between the decision maker/analyst and information source(s) that is developed in [31, 30, 32]. The latter can be thought of as a development of a general theory of inquiry that goes back to the work of Cox [6, 7]. This line of work received more attention recently resulting in a formulation of the calculus of inquiry [23] that constructs a distributive lattice of questions dual to the Boolean lattice of logical assertions. The definition of questions adapted in [31] corresponds to the particular subclass of questions – the partition questions – defined in [23]. Our work in [31, 30, 32] goes beyond that on the calculus of inquiry in that it introduces the concept of *pseudo-energy* as a measure of source specific difficulty of various questions to the given information source. One could say that it develops a quantitative theory of *knowledge* as opposed to the theory of information.

The problem of optimal usage of information obtained from experts has been addressed in existing research literature mostly in the form of updating the decision maker's beliefs given probability assessment from multiple experts [12, 13, 4, 5] and optimal combining of expert opinions, including experts with incoherent and missing outputs [34]. In the present and preceding papers, the emphasis is on *optimizing* on the particular type of information for the given expert(s) and decision making problem.

Methodologically, the present paper borrows heavily from the field of probability metrics and scenario reduction in stochastic optimization. More details, along with relevant references, can be found in Appendices.

The rest of the paper is organized as follows. In section 2, the description of the overall problem is given. Section 3 summarizes main facts and definitions concerning partitions of parameter spaces of decision making problems. In section 4, main results of [31, 30, 32] that are necessary for the developments in this paper are reviewed. Section 5 reviews main results of [33] where, in particular, the problem of additional information acquisition was formulated in the specific form that is used here. Section 6 develops the main theoretical framework for the use of scenario reduction methods for optimization of additional information acquisition. Section 7 develops specific algorithms for determining the efficient frontier and optimizing information acquisition. Section 8 provides an example illustrating the use of methods developed in section 7. Section 9 contains a conclusion. Appendices A and B review the necessary results in probability metrics and scenario reduction methods, respectively.

2 Problem Description

As was explained in [33], the starting point is the problem of the general form

$$\min_{x\in X}\mathbb{E}_P f(\omega,x)=\int_\Omega f(\omega,x)P(d\omega). \tag{1}$$

where X is the set of all feasible solutions, Ω is a parameter space to which uncertain problem parameters belong, and P is a fixed initial probability measure (with a suitable sigma-algebra assumed) on Ω that describes the initial state of the uncertainty. The function $f\colon\ \Omega\times X\to\overline{\mathbb{R}}$ is assumed to be integrable on Ω for each $x\in X$. For example, in the context of stochastic optimization, X is the set of feasible first-stage solutions and $f(\omega,x)$ is the best possible objective value for the first stage decision x in case when the random outcome ω is observed.

Let $L(P)$ be the expected loss corresponding to measure P defined as follows.

$$L(P)=\int_\Omega f(\omega,x^*_P)P(d\omega)-\int_\Omega f(\omega,x^*_\omega)P(d\omega), \tag{2}$$

where x^*_P is a solution of (1) and x^*_ω is a solution of $\min_{x\in X}f(\omega,x)$ for the given ω.

The main goal is, as explained in [33], for the given information source(s), find the way of extracting information from it so that the resulting expected loss would be minimized. The difference between the original loss (2 and the loss obtained with the help of the additional information extracted from the source can be termed the *value of information* following the definition proposed in [16]. Therefore, the main problem can be equivalently stated as maximization of value of information that can be provided by a given information source. The information can be extracted from the information source by means of asking question and receiving the source's answers. The decision maker can choose which specific questions to ask the given source so that the resulting value of information would take the maximum possible value.

Informally speaking, the problem is about finding the question(s) that is "aligned" optimally with both the information source's "strengths" and the particular decision making problem. Changing the purely "optimization" component of the problem (the function $f(\omega,x)$ and the set X) while keeping the "information" component (the space Ω and the measure P) the same will in general change the optimal question(s) for the same information source. Thus the main goal can also be described as that of finding an optimal alignment between the optimization and information components of the problem (where the information source itself is included in the latter).

3 Partitions of Parameter Space

In this section, we briefly summarize the necessary facts concerning partitions of the parameter space Ω which in general can be finite or infinite, such as a closed subset of a Euclidean space $\mathbb{R}^s$. We denote by $\mathcal{F}$ a sigma-algebra on Ω. Let P be a fixed probability measure on $(\Omega,\mathcal{F})$.

If $C\in\mathcal{F}$ be a measurable subset of Ω, We denote by P_C the conditional measure on Ω. It is defined in the usual way:

$$P_C(D)=\frac{P(D\cap C)}{P(C)}, \tag{3}$$

for arbitrary $D\in\mathcal{F}$.

A partition $\mathbf{C} = \{C_1, \ldots, C_r\}$ of Ω is a collection of measurable subsets $C_j \in \mathcal{F}$ of Ω such that $C_j \cap C_l = \emptyset$ for $j \neq l$ and $\cup_{j=1}^r C_j = \Omega$. A partition $\tilde{\mathbf{C}}$ is a *refinement* of $\mathbf{C}$ if every set from $\tilde{\mathbf{C}}$ is a subset of some set from $\mathbf{C}$. In such a case, $\mathbf{C}$ is a *coarsening* of $\tilde{\mathbf{C}}$. Given measure P on Ω, we call partition $\mathbf{C}_f(P)$ the *finest* partition of Ω associated with measure P if $P(C) > 0$ for all $C \in \mathbf{C}_f(P)$ and there exists at least one set of zero measure in any refinement of $\mathbf{C}_f(P)$. In case Ω is a closed subset of a Euclidean space and $\mathcal{F}$ is a Borel algebra, it is easy to see that finest partitions do not exist if measure P has a continuous support or has a component with continuous support. It is also clear that if the measure P has discrete support there exist many partitions of Ω that are finest for P.

Let $\mathbf{C}' = \{C'_1, \ldots, C'_r\}$ and $\mathbf{C}'' = \{C''_1, \ldots, C''_s\}$ be two partitions of Ω. Then partition $\mathbf{C} = \mathbf{C}' \cap \mathbf{C}''$ is defined as the partition that consists of all sets of the form $C'_i \cap C''_j$: $\mathbf{C}' \cap \mathbf{C}'' = \{C'_1 \cap C''_1, C'_1 \cap C''_2, \ldots, C'_r \cap C''_s\}$. Obviously, some of the sets constituting partition $\mathbf{C}' \cap \mathbf{C}''$ may be empty. Clearly, partition $\mathbf{C}' \cap \mathbf{C}''$ is a refinement of both $\mathbf{C}'$ and $\mathbf{C}''$.

If D is a subset of Ω and $\mathbf{C}' = \{C'_1, \ldots, C'_r\}$ is a partition of Ω, the partition $\mathbf{C}'_D = \{D \cap C'_1, \ldots, D \cap C'_r\}$ of D will be called the partition of D *induced* by the the partition $\mathbf{C}'$ of Ω.

For an arbitrary complete partition $\mathbf{C} = \{C_1, \ldots, C_r\}$, the measure P can be written as a linear combination of conditional measures:

$$P = \sum_{j=1}^{r} P(C_j) P_{C_j}, \tag{4}$$

which can be thought of as simply a generalization of an elementary total probability rule.

4 Information Exchange: Questions, Answers and Source Models

The model of information exchange between the decision maker/analyst and information source(s) was described in [31, 30, 32]. Here we review the main results to make the presentation self-contained. The main information exchange model contains the decision maker's questions, sources answers and the model of information source that relates the question difficulty to the answer depth.

4.1 Questions

Questions were identified in [31] with partitions $\mathbf{C} = \{C_1, \ldots, C_r\}$ of the parameter space Ω of the problem. Partitions were allowed to be incomplete, i.e. such that $\cup_{j=1}^r C_j \subset \Omega$. The *question difficulty* functional was introduced to measure the degree of difficulty of the question to the given information source, so that the information source would be able to answer questions with lower values of the difficulty functional more accurately that those with higher values of difficulty. The specific form of the difficulty functional was determined in [31] by demanding that it satisfy a system of reasonable postulates that, in particular, imposed the requirements of linearity and isotropy. The resulting form of the difficulty functional is given in the following theorem.

Theorem 1 *Let the function $G(\Omega, \mathbf{C}, P)$ where $\mathbf{C} = \{C_1, \ldots, C_r\}$ satisfy Postulates 1 through 6* *(see [31]). Then it has the form*

$$G(\Omega, \mathbf{C}, P) = \frac{\sum_{j=1}^{r} u(C_j) P(C_j) \log \frac{1}{P(C_j)}}{\sum_{j=1}^{r} P(C_j)},$$

where $u(C_j) = \frac{\int_{C_j} u(\omega)\, dP(\omega)}{P(C_j)}$ *and* $u\colon \Omega \to \mathbb{R}$ *is an integrable nonnegative function on the parameter space* Ω*.*

One can see that the difficulty of the given question $\mathbf{C}$ depends on, besides the initial probability measure P, the function $u : \Omega \to \mathbb{R}_+$ on the parameter space Ω. Using parallels with thermodynamics (see [31] for more details), this function may be called the *pseudo-temperature*. The question difficulty then can be interpreted as the amount of *pseudo-energy* associated with question $\mathbf{C}$.

4.2 Answers

Given a question $\mathbf{C} = \{C_1, \ldots, C_r\}$, the information source can provide an answer $V(\mathbf{C})$ that takes one of values in the set $\{s_1, \ldots, s_m\}$. A reception of the value s_k has an effect of modifying the original probability measure P on Ω to a new (updated) measure P^k. To ensure the the answer $V(\mathbf{C})$ is in fact an answer to the (complete) question $\mathbf{C}$ (and no more) the following condition is required to hold for the updated measures P^k, $k = 1, \ldots, m$:

$$P^k = \sum_{j=1}^{r} p_{kj} P_{C_j}, \tag{5}$$

where p_{kj}, $k = 1, \ldots, m$, $j = 1, \ldots, r$ are nonnegative coefficients such that $\sum_{j=1}^{r} p_{kj} = 1$ for $k = 1, \ldots, m$.

For incomplete questions, the expression (5) is modified to account for the set $\overline{\hat{C}} = \Omega \setminus \hat{C}$ and takes the form

$$P^k = \sum_{j=1}^{r} p_{kj} P_{C_j} + \bar{p}_k P_{\overline{\hat{C}}}, \tag{6}$$

where $\sum_{j=1}^{r} p_{kj} + \bar{p}_k = 1$.

The *answer depth* functional $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ for the answer $V(\mathbf{C})$ to question $\mathbf{C}$ measures the amount of *pseudo-energy* that is conveyed by $V(\mathbf{C})$ in response to question $\mathbf{C}$. The general form of $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ can be established if certain reasonable requirements (postulates) it has to satisfy are imposed. A system of postulates proposed in [30] that parallels the postulates for question difficulty and, in particular, imposes the requirements of linearity and isotropy. The following theorem was then proved in [30].

Theorem 2 *The answer depth functional* $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ *has the form*

$$Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) = \sum_{k=1}^{m} \Pr(V(\mathbf{C}) = s_k) \frac{\sum_{j=1}^{r} u(C_j) P^k(C_j) \log \frac{P^k(C_j)}{P(C_j)}}{\sum_{j=1}^{r} P^k(C_j)},$$

where $P^k \equiv P^{V(\mathbf{C})=s_k}$ *is the measure on* Ω *conditioned on reception of* $V(\mathbf{C}) = s_k$ *and* $u(C_j) = \frac{1}{P(C_j)} \int_{C_j} u(\omega) dP(\omega)$ *and the function* $u \colon \Omega \to \mathbb{R}$ *is the same function that is used in the question difficulty functional* $G(\Omega, \mathbf{C}, P)$.

It can be shown (see [30] for details) that if $V(\mathbf{C})$ is any answer to the question $\mathbf{C}$ then $Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) \leq G(\Omega, \mathbf{C}, P)$ with equality if and only if the answer $V(\mathbf{C}) = V^*(\mathbf{C})$ is *perfect*, i.e. $P^j = P_{C_j}$ for $j = 1, \ldots, r$.

As far as answers that are not perfect are concerned, it is convenient to consider the class of answers for which the degree of imperfection is described by a single error probability α – the *quasi-perfect* answers [30]. For a quasi-perfect answer $V_\alpha(\mathbf{C})$ to a (complete) question $\mathbf{C} = \{C_1, \ldots, C_r\}$, the coefficients p_{kj} have the form

$$p_{kj} = (1 - \alpha)\delta_{k,j} + \alpha P(C_j), \tag{7}$$

for $k = 1, \ldots, r$ and $j = 1, \ldots, r$, and the updated measure P^k is simply

$$P^k = \alpha P + (1 - \alpha) P_{C_k}. \tag{8}$$

for $k = 1, \ldots, r$. Clearly, for $\alpha = 0$ a quasi-perfect answer to $\mathbf{C}$ becomes a perfect one. It can be shown (see [30]) that the answer depth functional for a quasi-perfect answer $V_\alpha(\mathbf{C})$ to question $\mathbf{C}$ can be written as

$$\begin{aligned} Y(\Omega, \mathbf{C}, P, V_\alpha(\mathbf{C})) = &\sum_{k=1}^{r} u(C_k) P(C_k)(1 - \alpha + \alpha P(C_k)) \log \frac{1 - \alpha + \alpha P(C_k)}{P(C_k)} \\ &+ \alpha \log \alpha \sum_{k=1}^{r} u(C_k) P(C_k)(1 - P(C_k)), \end{aligned} \tag{9}$$

which can be seen to reduce to $G(\Omega, \mathbf{C}, P)$ for $\alpha = 0$ (when $V(\mathbf{C}) = V^*(\mathbf{C})$) and vanish for $\alpha = 1$.

4.3 Information Source Models

If the pseudo-temperature function $u(\cdot)$ for the given information source is known, the difficulty of any question to the source can be determined as stated in Theorem 1. On the other hand, given the updated measures P^k corresponding to all possible values of answer $V(\mathbf{C})$, the depth of the answer can be found from Theorem 2. What's lacking in this picture is the information source characteristic: what is the depth of an answer it can provide in response to a question of given difficulty. This gives rise to information source models discussed in [32].

An information source model, as it was defined in [32] is simply a (non-decreasing) function $h : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) = h(G(\Omega, \mathbf{C}, P)).$$

It also makes sense to assume that the function h is bounded from above by a value that can be identified with the source's *capacity* that has the meaning of a highest answer depth the source is capable of.

The simplest information source model considered in [32] is the so called *simple capacity model* given by

$$h(x) = \begin{cases} x & \text{if } x \leq Y_s \\ Y_s & \text{if } x > Y_s. \end{cases} \tag{10}$$

which is fully characterized by the single parameter Y_s which has the meaning of the information source capacity.

The most apparent drawback of model (10) is that, according to it, the source provides a perfect answer to any question with difficulty not exceeding the source capacity. The *linear modified capacity model* described by

$$h(x) = \begin{cases} bx & \text{if } x \leq \frac{Y_s}{b} \\ Y_s & \text{if } x > \frac{Y_s}{b} \end{cases} \tag{11}$$

removes this drawback at the expense of one extra parameter $b \leq 1$ that has to be estimated. Several other models were proposed in [32].

The values of model parameters as well as pseudo-temperature functions for information sources can be estimated from the observed sources' performance on sample questions. The corresponding estimation procedures were also discussed in [32].

5 Value of Information: Loss Reduction

To describe the value of information for the given problem of the form (1) and the problem of additional information acquisition optimization we need to introduce maps from Ω into X and some functionals of such maps and probability measures on Ω.

Let $\mathcal{G}$ be the set of all such maps with a discrete image set. Clearly, any such map $g \in \mathcal{G}$ can be uniquely described by the corresponding partition $\mathbf{C} = \{C_1, \dots, C_r\}$ of Ω and the corresponding image set $I = \{x_1, \dots, x_r\}$ such that $g(\omega) = x_j$ for all $\omega \in C_j$.

Let P be any probability measure on Ω, let x an arbitrary element of the solution space X, and let $g \in \mathcal{G}$ be an arbitrary map from Ω into X. The *suboptimality*, *loss* and *gain* functionals are defined ([33]) as follows.

$$S(x, P) = \mathbb{E}_P f(\omega, x) - \mathbb{E}_P f(\omega, x^*_P) = \int_\Omega (f(\omega, x) - f(\omega, x^*_P))P(d\omega), \tag{12}$$

$$L(g, P) = \mathbb{E}_P f(\omega, g(\omega)) - \mathbb{E}_P f(\omega, x^*_\omega) = \int_\Omega (f(\omega, g(\omega)) - f(\omega, x^*_\omega))P(d\omega), \tag{13}$$

and

$$B(g, P) = \mathbb{E}_P f(\omega, x^*_P) - \mathbb{E}_P f(\omega, g(\omega)) = \int_\Omega (f(\omega, x^*_P) - f(\omega, g(\omega)))P(d\omega). \tag{14}$$

respectively.

Moreover, it is convenient to introduce the corresponding functionals not just for a fixed measure P, but also for the given question $\mathbf{C}$ and a given answer $V(\mathbf{C})$, For example, for an arbitrary $x \in X$, the suboptimality of solution x with respect to question $\mathbf{C}$ (and initial measure P) is given by

$$S(x, P_\mathbf{C}) = \sum_{i=1}^{s} P(C_j) S(x, P_{C_j}), \tag{15}$$

and the suboptimality of x with respect to answer $V(\mathbf{C})$ to question $\mathbf{C}$ (and initial measure P) reads

$$S(x, P_{V(\mathbf{C})}) = \sum_{k=1}^{m} v_k S(x, P^k), \tag{16}$$

where $v_k \equiv \Pr(V(\mathbf{C}) = s_k)$ for brevity. The loss and gain functionals for the given map $g \in \mathcal{G}$ and question $\mathbf{C}$ and answer $V(\mathbf{C})$ are defined analogously.

Note that each map $g = (\mathbf{C}(g), I(g))$ from the set $\mathcal{G}$ can be characterized by the corresponding loss $L(g, P)$ with respect to the original measure P and the value $G(\Omega, \mathbf{C}(g), P)$ – the difficulty of the corresponding question. The *efficient frontier* in the Euclidean plane with coordinates $(G(\Omega, \mathbf{C}(g), P), L(g, P))$ can be found by solving the following parametric optimization problem

$$\begin{aligned} \underset{g \in \mathcal{G}}{\text{minimize}} \quad & L(g, P) \\ \text{subject to} \quad & G(\Omega, \mathbf{C}(g), P) \leq \gamma \end{aligned} \tag{17}$$

for all values of the parameter γ.

The maps g that are solutions of (17) for various values of the parameter γ have the property of having the smallest possible loss among the maps corresponding to question whose difficulty does not exceed the given value γ. Let us denotes by $\mathcal{O}$ the set of all maps in $\mathcal{G}$ that are solutions of (17) and by $\mathcal{C}$ the set of all *subset-optimal* maps, i.e. maps of the form $(\{C_1, \ldots, C_r\}, \{x^*_{P_{C_1}}, \ldots, x^*_{P_{C_r}}\})$, where $x^*_{P_{C_j}}$ is an optimal solution of problem (1) with measure P replaced with the conditional measure P_{C_j}. Then, as was shown in [33],

$$\mathcal{O} \subseteq \mathcal{C}, \tag{18}$$

i.e. if one is interested in finding Pareto-optimal maps in $\mathcal{O}$ it is sufficient to consider subset-optimal maps only.

Let us now address the optimal information acquisition problem (20): what question(s) need to be asked the given information source in order to obtain the minimum possible loss for (1). Given a question $\mathbf{C} = \{C_1, \ldots, C_r\}$ to an information source and its answer $V(\mathbf{C})$ taking values in the set $\{s_1, \ldots, s_m\}$, we denote by $\mathcal{L}(s_k)$, $k = 1, \ldots, m$ the *minimum conditional expected loss* given that $V(\mathbf{C}) = s_k$ and by $\mathcal{L}(V(\mathbf{C}))$ the *minimum expected loss* that the decision maker can achieve given the answer $V(\mathbf{C})$. The latter can be found as

$$\mathcal{L}(V(\mathbf{C})) = \sum_{k=1}^{m} \Pr(V(\mathbf{C}) = s_k)\mathcal{L}(s_k), \tag{19}$$

i.e. as an expectation over possible values of the answer $V(\mathbf{C})$.

If the decision maker poses a question $\mathbf{C} = \{C_1, \ldots, C_r\}$ to the information source and receives a particular value s_k of answer $V(\mathbf{C})$, the original measure P on Ω gets updated to $P^k \equiv P^{V(\mathbf{C})=s_k}$. Therefore in order to minimize loss for the given value s_k of answer $V(\mathbf{C})$ the decision maker needs to choose the solution $x^*_{P^k}$ – the solution minimizing the expectation $\mathbb{E}_{P^k} f(\omega, x)$ over all (feasible) values of x.

The next two propositions, proved in [33], give the minimum expected loss achievable with a perfect and a general answer to question $\mathbf{C}$, respectively.

Proposition 1 *Let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be a complete question and $g_{\mathbf{C},P} \in \mathcal{C}$ be a corresponding subset optimal map. If the decision maker is given a perfect answer $V^*(\mathbf{C})$ to $\mathbf{C}$ then*

$$\mathcal{L}(V^*(\mathbf{C})) = L(g_{\mathbf{C},P}, P).$$

Proposition 2 *Let* $\mathbf{C} = \{C_1, \ldots, C_r\}$ *be a complete question and* $g_{\mathbf{C},P} \in \mathcal{C}$ *be a corresponding subset optimal map. If the decision maker is given a (generally imperfect) answer* $V(\mathbf{C})$ *to* $\mathbf{C}$ *then*

$$\mathcal{L}(V(\mathbf{C})) = B(g_{\mathbf{C},P}, P_{V(\mathbf{C})}) + L(g_{\mathbf{C},P}, P).$$

The information acquisition optimization problem can then be written as

$$\begin{aligned} &\underset{\mathbf{C}}{\text{minimize}} && \mathcal{L}(V(\mathbf{C})) \\ &\text{subject to} && Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) = h(G(\Omega, \mathbf{C}, P)), \end{aligned} \tag{20}$$

where the minimum expected loss $\mathcal{L}(V(\mathbf{C}))$ is given by either Proposition 1 or Proposition 2. The source model function $h(\cdot)$ and the pseudo-temperature function $u(\cdot)$ that enters the expressions for the question difficulty and answer depth in (20) are assumed to be known.

It's easy to see that if a source is capable of perfect answers (for instance, in the simple linear model) solution of problem (20) reduces to finding the efficient frontier: if $L^*(G)$ is the expression describing the efficient frontier (abstracting from its true discrete structure) and Y_s is the capacity of the information source, then the minimum in (20) is equal to $L^*(Y_s)$ and is achieved by the question $\mathbf{C}$ lying on the efficient frontier such that $G(\Omega, \mathbf{C}, P) = Y_s$.

If a source cannot provide perfect answers, questions with difficulty exceeding the source capacity need to be considered in order to minimize the expected loss. The search for an optimal question in this case becomes more complicated as the error structure for the source's answers needs to be taken into account. If answers are assumed to be quasi-perfect, optimal question(s) can be found approximately provided the efficient frontier is already known.

6 Information Acquisition Optimization

In the following, we assume that the (initial) probability measure P is supported at a discrete set $\{\omega_1, \ldots, \omega_N\} \equiv \Omega_N \subset \Omega$:

$$P = \sum_{i=1}^{N} p_i \delta_{\omega_i}, \tag{21}$$

where δ_ω is a Dirac delta that puts a unit mass at ω. Points $\omega_i \in \Omega_N$ are usually referred to as *scenarios*. The *scenario reduction* methodology (see Appendix) is often used in stochastic optimization to lower computational complexity of various practically important problems. In scenario reduction approach, the original discrete measure P given by (21) is said to be *reduced* to another discrete measure Q given by

$$Q = \sum_{j=1}^{M} q_j \delta_{\tilde{\omega}_j}, \tag{22}$$

if the support $\{\tilde{\omega}_1, \ldots, \tilde{\omega}_M\}$ of Q is a subset of Ω_N.

For later convenience, we denote by $\mathcal{R}_M(\Omega_N)$ the set of all *scenario reduction maps* from the set of measures of the form (21) supported at Ω_N into the set of all measures of the form (22) supported at some subset of Ω_N of cardinality $M < N$ satisfying the additional property that we call *simplicity.* A map $\nu \in \mathcal{R}_M(\Omega_N)$ is called *simple* if there exists a partition $\{S_1, \dots, S_M\}$ of the set of scenarios Ω_N such that $\nu(\omega_i) = \tilde{\omega}_j$ for all $\omega_i \in S_j$ and $q_j = \sum_{\{i:\omega_i \in S_j\}} p_i$. In such a case we write $Q = \nu(P)$ and $S_j = \nu^{-1}(\tilde{\omega}_j)$ for $j = 1, \dots, M$.

Additionally, if c: $\Omega \times \Omega \to \mathbb{R}_+$ is some symmetric cost function, we call a map $\nu \in \mathcal{R}_M(\Omega_N)$ *c-optimal* if $i = \arg\min c(\omega_i, \nu(o_i))$ for $i = 1, \dots, N$. It is shown in [15] that the Monge-Kantorovich functional (see Appendix A) $\hat{\mu}_c(P, Q)$ is minimized for all measures Q supported at $\{\tilde{\omega}_1, \dots, \tilde{\omega}_M\} = \nu(\Omega_N)$ iff the corresponding simple scenario reduction map is c-optimal.

In the following we call measures P and Q **C**-*equivalent* for some partition $\mathbf{C}$ of Ω if $P(C) = Q(C)$ for all $C \in \mathbf{C}$. It is easy to see that measures P and Q are **C**-equivalent for all possible partitions $\mathbf{C}$ if and only if $P = Q$ but two distinct measures can easily be **C**-equivalent for a specific partition $\mathbf{C}$. In particular, any two measures on Ω are **C**-equivalent if $\mathbf{C}$ is the trivial partition $\mathbf{C} = \{\Omega\}$.

Given measure P on Ω and some measure Q that was obtained from P by a reduction, let us denote by $\mathcal{Q}(Q|P)$ the *virtual pseudo-energy* content of measure Q relative to P. It is defined as follows.

$$\mathcal{Q}(Q|P) = G(\Omega, \mathbf{C}_f(P), P) - G(\Omega, \mathbf{C}_f(Q), Q), \tag{23}$$

i.e. $\mathcal{Q}(Q|P)$ is the difference between the difficulties of exhaustive questions associated with measures P and Q, respectively. One can think about the virtual pseudo-energy of Q relative to P as an amount pseudo-energy a source would need to supply in order to obtain a new state in which the hardest possible question has a difficulty equal to $G(\Omega, \mathbf{C}_f(Q), Q)$. Since no question is in fact answered in going from measure P to the reduced measure Q we call this pseudo-energy virtual.

We can now introduce the *virtual difficulty* of question $\mathbf{C}$ for measure Q with respect to measure P:

$$G_P(\Omega, \mathbf{C}, Q) = \mathcal{Q}(Q|P) + G(\Omega, \mathbf{C}, Q). \tag{24}$$

In particular, $G_P(\Omega, \mathbf{C}, P) = G(\Omega, \mathbf{C}, P)$, i.e. the virtual difficulty of $\mathbf{C}$ for measure P relative to P reduces just to the standard difficulty of $\mathbf{C}$.

It also turns out to be useful to introduce the *relative expected loss* for partitions of Ω and measures Q obtained from the original measure P by a (simple) scenario reduction operation. In other words, we assume that there exists $\nu \in \mathcal{R}_M(\Omega_N)$ for some value of $M < N$ such that $Q = \nu(P)$. The relative (to measure P) expected loss of partition $\mathbf{C}$ and measure Q is then defined as follows.

$$L_P(\mathbf{C}, Q) = \sum_{C \in \mathbf{C}} P(C) L(g_{\mathbf{C},Q}, P), \tag{25}$$

where $g_{\mathbf{C},Q}$ is the subset-optimal map for partition $\mathbf{C}$ and measure Q. In particular, if $\mathbf{C}$ is the trivial partition $\mathbf{C} = \{\Omega\}$, the loss of Q relative to P is simply[1] $L_P(Q) = L(g_Q, P)$. If the measure Q coincides with P, the loss relative to P is just the standard expected loss of the corresponding subset-optimal map: $L_P(\mathbf{C}, P) = L(g_{\mathbf{C},P}, P)$.

Let us now consider the following construction. Reduce the original measure P to Q that is supported at r points: $Q = \nu(P)$, where $\nu \in \mathcal{R}_r(\Omega_N)$. Let $Q = \sum_{j=1}^{r} q_j \tilde{\omega}_j$ and let S_j the preimage of $\tilde{\omega}_j$ under map ν: $\nu(\omega_i) = \tilde{\omega}_j$ for all $\omega_i \in S_j$. Then let $\mathbf{C}$ be a partition of Ω such that $S_j \subset C_j$ for $j = 1, \dots, r$. We say that the partition $\mathbf{C}$ is *generated* by the map $\nu \in \mathcal{R}_r(\Omega_N)$, or, equivalently by the reduction of measure P to Q. Let $\hat{\mathbf{C}}$ be an arbitrary coarsening of $\mathbf{C}$.

[1]Here and later we omit the trivial partition from the list of arguments of $G(\cdot)$ and $L(\cdot)$.

We are interested in the location of points P, Q, $(\mathbf{C}, P)$, $(\mathbf{C}, Q)$, $(\hat{\mathbf{C}}, P)$ and $(\hat{\mathbf{C}}, Q)$ on the plane with coordinates $(G_P(\Omega, \cdot), L_P(\cdot))$. First of all, it is clear that $G_P(\Omega, P) = 0$ and $L_P(P) = L(g_P, P)$ where $L(g_P, P)$ is the EVPI of problem (1). Second, it is also clear that

$$\begin{aligned} G_P(\Omega, \mathbf{C}, Q) &= \mathcal{Q}(Q|P) + G(\Omega, \mathbf{C}, Q) \\ &= G(\Omega, \mathbf{C}_f(P), P) - G(\Omega, \mathbf{C}_f(Q), Q) + G(\Omega, \mathbf{C}, Q) \\ &= G(\Omega, \mathbf{C}_f(P), P) \end{aligned} \tag{26}$$

since $\mathbf{C} = \mathbf{C}_f(Q)$ by construction of Q. In words, the virtual difficulty of the question $\mathbf{C}$ for measure Q where the partition $\mathbf{C}$ was generated by a reduction of the original measure P to Q is equal to the difficulty of the exhaustive question for the original measure P.

To obtain relationships between relative expected losses the following two auxiliary lemmas are needed.

Lemma 1 *Let $c_{ij} = c_{ji}$, $i, j = 1, \ldots, N$ be a symmetric matrix with elements c_{ij} satisfying the triangle inequality $c_{ij} \leq c_{ik} + c_{kj}$. Let $\{p_i\}_{i=1}^N$ be a probability distribution. Then*

$$\sum_{i=1}^N \sum_{j=1}^N p_j p_j c_{ij} \leq 2 \min_i \sum_{j=1}^N p_j c_{ij}.$$

Proof: Let $i^* = \arg\min_i \sum_{j=1}^N p_j c_{ij}$ (so that $\min_i \sum_{j=1}^N p_j c_{ij} = \sum_{j=1}^N p_j c_{i^*j}$). Then we can write

$$\begin{aligned} \sum_{i=1}^N \sum_{j=1}^N p_i p_j c_{ij} &\overset{(a)}{\leq} \sum_{i=1}^N \sum_{j=1}^N p_i p_j (c_{ii^*} + c_{i^*j}) \\ &= \sum_{i=1}^N \sum_{j=1}^N p_i p_j c_{ii^*} + \sum_{i=1}^N \sum_{j=1}^N p_i p_j c_{i^*j} \\ &= \sum_{j=1}^N p_j \sum_{i=1}^N p_i c_{i^*i} + \sum_{i=1}^N p_i \sum_{j=1}^N p_j c_{i^*j} \\ &\overset{(b)}{=} 2 \min_i \sum_{j=1}^N p_j c_{ij}, \end{aligned}$$

where (a) follows from the triangle inequality satisfied by the elements c_{ij} and (b) follows from the definition of i^*. □

The second lemma states a useful probability metrics result. Let $P = \sum_{i=1}^N p_i \delta_{\omega_i}$ be a discrete support probability measure on Ω and let $Q = \sum_{i=1}^M q_i \delta_{\tilde{\omega}_i}$ be another such measure. Let $\zeta_c(P, Q)$ be a Fortet-Mourier metric for some cost function $c : \Omega \times \Omega \to \mathbb{R}_+$ that satisfies conditions described in Appendix 1. Finally, let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be a partition of Ω such that the measures P and Q are $\mathbf{C}$-equivalent.

Lemma 2 *Under assumptions described above,*

1. *$\zeta_c(P, Q) \leq \sum_{j=1}^r w_j \zeta_c(P_{C_j}, Q_{C_j})$, where $w_j = P(C_j) = Q(C_j)$.*

2. If Q is generated by some map $\nu \in \mathcal{R}_r(\Omega_N)$ that is $\hat{c}$-optimal, where $\hat{c}$ is the reduced cost function defined as in (36) then

$\zeta_c(P, Q) = \sum_{j=1}^{r} w_j \zeta_c(P_{C_j}, Q_{C_j})$.

Proof: The first statement actually holds true for any measures $P, Q \in \mathcal{P}_c(\Omega)$ (see Appendix 1 for the definition of $\mathcal{P}_c(\Omega)$). Indeed, let $f^*(\omega) \in \mathcal{F}_c$ be the function that achieves the maximum of

$$\left| \int_{\Omega} f(\omega)P(d\omega) - \int_{\Omega} f(\omega)Q(d\omega) \right|.$$

Let $f_j^*(\omega)$ be the restriction of $f^*(\omega)$ to C_j. Clearly, $f_j^*(\omega) \in \mathcal{F}_c(C_j)$. We can write

$$\zeta_c(P, Q) = \left| \int_{\Omega} f^*(\omega)P(d\omega) - \int_{\Omega} f(\omega)Q(d\omega) \right| \overset{(a)}{=} \sum_{j=1}^{r} w_j \left| \int_{C_j} f^*(\omega)dP_{C_j}(\omega) - \int_{C_j} f^*(\omega)dQ_{C_j}(\omega) \right|$$
$$\overset{(b)}{=} \sum_{j=1}^{r} w_j \left| \int_{C_j} f_j^*(\omega)dP_{C_j}(\omega) - \int_{C_j} f_j^*(\omega)dQ_{C_j}(\omega) \right| \overset{(c)}{\leq} \sum_{j=1}^{r} w_j \zeta_c(P_{C_j}, Q_{C_j}),$$

where (a) follows from the definition of conditional measures P_{C_j} and Q_{C_j}, (b) follows from the definition of functions $f_j^*(\omega)$, and (c) follows from that $f_j^*(\omega) \in \mathcal{F}_c(C_j)$ and definition of $\zeta_c(P_{C_j}, Q_{C_j})$.

To prove the second statement, we can use the duality result (33) described in Appendix 1 together with (38) that relates the values of Kantorovich-Rubinstein and Monge-Kantorovich functionals. Let $\nu \in \mathcal{R}_r(\Omega_N)$ be the map that generates partition $\mathbf{C}$, and let $\tilde{\omega}_j = \nu(\omega_i)$ for all $\omega_i \in \mathbf{C}_j$. Note also that $q_j = \sum_{\{i:\omega_i \in C_j\}} p_i = w_j$, $j = 1, \ldots, r$. We can write

$$\sum_{j=1}^{r} w_j \zeta_c(P_{C_j}, Q_{C_j}) \overset{(a)}{=} \sum_{j=1}^{r} w_j \hat{\mu}_{\hat{c}}(P_{C_j}, Q_{C_j}) \overset{(b)}{=} \sum_{j=1}^{r} w_j \sum_{\{i:\omega_i \in C_j\}} \frac{p_i}{w_j} \hat{c}(\omega_i, \tilde{\omega}_j)$$
$$= \sum_{j=1}^{r} \sum_{\{i:\omega_i \in C_j\}} p_i \hat{c}(\omega_i, \tilde{\omega}_j) \overset{(c)}{=} \hat{\mu}_{\hat{c}}(P, Q) \overset{(d)}{=} \zeta_c(P, Q),$$

where (a) and (d) follow from (33) and (38), (b) follows from that Q_{C_j} is supported at a single point $\tilde{\omega}_j$, (c) follows from the way measure Q was constructed as a reduction of the measure P with a $\hat{c}$-optimal map $\nu \in \mathcal{R}_r(\Omega_N)$. □

Now, assume that the integrand $f(\omega, x)$ in (1) is in class $\mathcal{F}_c$ defined in Appendix A (expression (31)) for some symmetric cost function $c : \Omega \times \Omega \to \mathbb{R}_+$ that satisfies the conditions described in Appendix A. The following proposition describes a relation between relative expected losses for measures P and Q.

Proposition 3 *Let $\mathbf{C}$ be a partition of Ω generated by a reduction of a measure P with support at $\Omega_N \subset \Omega$ to Q by means of a $\hat{c}$-optimal map $\nu \in \mathcal{R}_r(\Omega_N)$ and let $\hat{\mathbf{C}}$ any coarsening of $\mathbf{C}$ (including $\mathbf{C}$ itself). Then*

$$L_P(\hat{\mathbf{C}}, Q) \leq L_P(\hat{\mathbf{C}}, P) + 2K\zeta_c(P, Q),$$

where $K > 0$ is some constant that does not depend on measures Q and P.

Proof: Let $w_j = P(\hat{C}_j) = Q(\hat{C}_j)$ be the measure of subsets in $\hat{\mathbf{C}}$ and let $P_j \equiv P_{\hat{C}_j}$ and $Q_j \equiv Q_{\hat{C}_j}$ be the corresponding subset measures.

$$
\begin{aligned}
L_P(\hat{\mathbf{C}}, Q) &= \sum_{j=1}^{r} w_j \left[\int_{\hat{C}_j} f(\omega, x^*_{Q_j}) P_j(d\omega) - \int_{\hat{C}_j} f(\omega, x^*_{\omega}) P_j(d\omega) \right] \\
&= \sum_{j=1}^{r} w_j \left[\int_{\hat{C}_j} f(\omega, x^*_{Q_j}) P_j(d\omega) - \int_{\hat{C}_j} f(\omega, x^*_{\omega}) P_j(d\omega) + \int_{\hat{C}_j} f(\omega, x^*_{P_j}) P_j(d\omega) - \int_{\hat{C}_j} f(\omega, x^*_{P_j}) P_j(d\omega) \right] \\
&\overset{(a)}{=} L_P(\hat{\mathbf{C}}, P) + \sum_{j=1}^{r} w_j \left[\int_{\hat{C}_j} f(\omega, x^*_{Q_j}) P_j(d\omega) - \int_{\hat{C}_j} f(\omega, x^*_{P_j}) P_j(d\omega) \right] \\
&= L_P(\hat{\mathbf{C}}, P) + \sum_{j=1}^{r} w_j \left[\int_{\hat{C}_j} f(\omega, x^*_{Q_j}) P_j(d\omega) - \int_{\hat{C}_j} f(\omega, x^*_{P_j}) P_j(d\omega) \right. \\
&\qquad \left. + \int_{\hat{C}_j} f(\omega, x^*_{Q_j}) Q_j(d\omega) - \int_{\hat{C}_j} f(\omega, x^*_{Q_j}) Q_j(d\omega) \right] \\
&\overset{(b)}{=} L_P(\hat{\mathbf{C}}, P) + \sum_{j=1}^{r} w_j \left[v(Q_j) - v(P_j) + \int_{\hat{C}_j} f(\omega, x^*_{Q_j})(P_j - Q_j)(d\omega) \right] \\
&\leq L_P(\hat{\mathbf{C}}, P) + \sum_{j=1}^{r} w_j \left| v(Q_j) - v(P_j) \right| + \sum_{j=1}^{r} w_j \left| \int_{\hat{C}_j} f(\omega, x^*_{Q_j})(P_j - Q_j)(d\omega) \right| \\
&\overset{(c)}{\leq} L_P(\hat{\mathbf{C}}, P) + K \sum_{j=1}^{r} w_j \zeta_c(P_j, Q_j) + K \sum_{j=1}^{r} w_j \zeta_c(P_j, Q_j) = L_P(\hat{\mathbf{C}}, P) + 2K \sum_{j=1}^{r} w_j \zeta_c(P_j, Q_j) \\
&\overset{(d)}{=} L_P(\hat{\mathbf{C}}, P) + 2K \zeta_c(P, Q)
\end{aligned}
$$

where (a) follows from the definition of $L_P(\hat{\mathbf{C}}, P)$, (b) follows from the definition of the optimal objective values $v(P_j)$ and $v(Q_j)$, (c) follows from that the integrand $f(\omega, x)$ is in $\mathfrak{F}_c$ and definition (32) of Fortet-Mourier metric ζ_c, and (d) follows from Lemma 2.□

If we use the trivial partition $\hat{\mathbf{C}} = \{\Omega\}$ (which is obviously a coarsening of any $\mathbf{C}$) in Proposition 3 we can obtain an upper bound on the relative loss of Q with respect to P which we formulate as a corollary.

Corollary 1 *The loss of reduced measure Q relative to P can be bounded from above as*

$$L_P(Q) \leq L(g_P, P) + 2K\zeta_c(P, Q),$$

where $L(g_P, P) \equiv L_P(P)$ is the EVPI of the original problem (1).

The following proposition relates the expected loss of a subset-optimal map based on a partition generated by a reduction of the original measure P to measure Q to the Fortet-Mourier distance between P and Q.

Proposition 4 *Let $\mathbf{C}$ be a partition of Ω generated by a reduction of a measure P supported at the discrete set $\Omega_N \subset \Omega$ to measure Q by means of a $\hat{c}$-optimal map $\nu \in \mathfrak{F}_r(\Omega_N)$. Then*

$$L_P(\mathbf{C}, P) \equiv L(g_{\mathbf{C},P}, P) \leq 2K\zeta_c(P, Q),$$

where $K > 0$ is a constant.

Proof: Let $w_j = P(C_j) = P(Q_j)$, $j = 1, \ldots, r$ be measures of subsets in $\mathbf{C}$ and let P_j and Q_j be the corresponding subset measures.

$$
\begin{aligned}
L(g_{\mathbf{C},P}, P) &= \sum_{j=1}^{r} w_j L(g_{P_j}, P_j) = \sum_{j=1}^{r} w_j \int_{C_j} \Big(f(\omega, x^*_{P_j}) - f(\omega, x^*_{\omega}) \Big) P_j(d\omega) \\
&= \sum_{j=1}^{r} w_j \sum_{\{i:\omega_i \in C_j\}} \frac{p_i}{w_j} \Big(f(\omega_i, x^*_{P_j}) - f(\omega_i, x^*_{\omega_i}) \Big) \\
&\overset{(a)}{=} \sum_{j=1}^{r} w_j \left(v(P_j) - \sum_{\{i:\omega_i \in C_j\}} (P_j)_i \, v(\delta_{\omega_i}) \right) = \sum_{j=1}^{r} w_j \sum_{\{i:\omega_i \in C_j\}} (P_j)_i \left(v(P_j) - v(\delta_{\omega_i}) \right) \\
&\overset{(b)}{\le} K \sum_{j=1}^{r} w_j \sum_{\{i:\omega_i \in C_j\}} (P_j)_i \, \zeta_c(P_j, \delta_{\omega_i}) \overset{(c)}{=} K \sum_{j=1}^{r} w_j \sum_{\{i:\omega_i \in C_j\}} (P_j)_i \, \hat{\mu}_{\hat{c}}(P_j, \delta_{\omega_i}) \\
&= K \sum_{j=1}^{r} w_j \sum_{\{i:\omega_i \in C_j\}} (P_j)_i \sum_{\{k:\omega_k \in C_j\}} (P_j)_k \, \hat{c}(\omega_i, \omega_k) \\
&\overset{(d)}{\le} 2K \sum_{j=1}^{r} w_j \min_{\{k:\omega_k \in C_j\}} \sum_{\{i:\omega_i \in C_j\}} (P_j)_i \, \hat{c}(\omega_i, \omega_k) = 2K \sum_{j=1}^{r} w_j \min_{\{k:\omega_k \in C_j\}} \hat{\mu}_{\hat{c}}(P_j, \delta_{\omega_k}) \\
&\overset{(e)}{=} 2K \sum_{j=1}^{r} w_j \hat{\mu}_{\hat{c}}(P_j, Q_j) = 2K \sum_{j=1}^{r} w_j \zeta_c(P_j, Q_j) \overset{(f)}{=} 2K \zeta_c(P, Q),
\end{aligned}
$$

where $(P_j)_i \equiv \frac{p_i}{w_j}$ for $\omega_i \in C_j$, (a) follows from the definition of optimal values $v(P_j)$ and $v(\delta_{\omega_i})$, (b) follows from the upper bound (39), (c) follows from the duality relation (33) and from the relation (38) between the Kantorovich-Rubinstein and Monge-Kantorovich functionals, (d) follows from Lemma 1 (since $\hat{c}$ is a metric and $\{(P_j)_i\}_{\{i:\omega_i \in C_j\}}$ is a probability distribution), (e) follows from that $Q = \nu(P)$, where ν is $\hat{c}$-optimal, and (f) follows from Lemma 2.□

Fig. 1 shows the locations of various points on $(G_P(\Omega, \cdot, \cdot), L_P(\cdot, \cdot))$ coordinate plane.

Several useful observations can now be made.

- The result of Proposition 4 suggests that good (near-optimal) partitions of Ω can be generated by a reduction of the original measure P to a measure Q that is (i) supported at a few points and (ii) has a low value of the Fortet-Mourier metric $\zeta_c(P, Q) = \hat{\mu}_{\hat{c}}(P, Q)$. The latter value of the Monge-Kantorovich functional $\hat{\mu}_{\hat{c}}(P, Q)$ with the reduced cost $\hat{c}$ can be readily computed as that of a minimum-cost transportation problem.

- For a wide class of linear multi-period two stage stochastic optimization problems, the relevant cost function c is given by c_p (see Appendix A, expression (40)) with $p = l + 1$ where l is the number of periods. The corresponding minimum cost transportation problem can be easily solved exactly for fixed support of measure Q and approximately if the support itself needs to be optimized (see Appendix B for details).

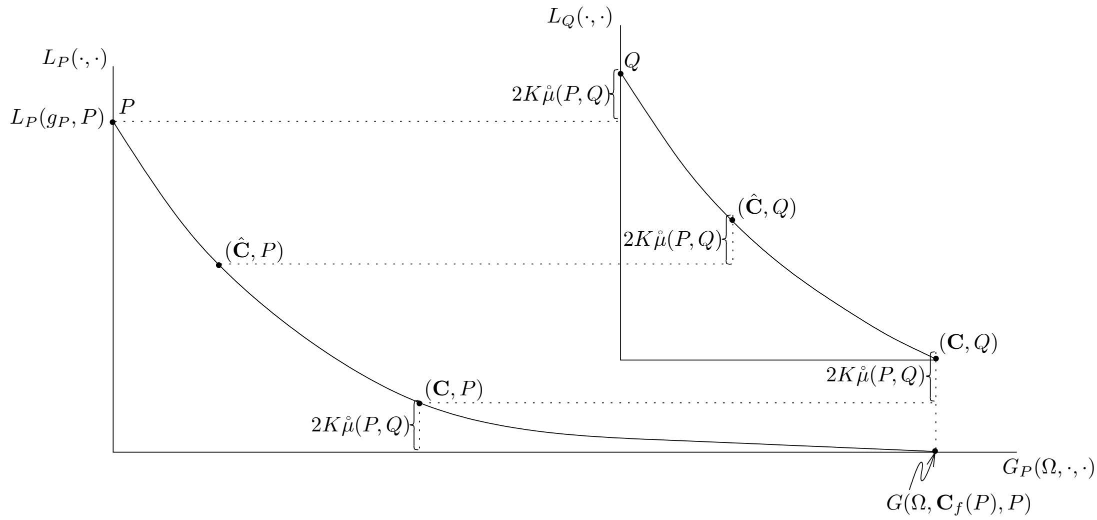

Figure 1: Pseudo-energy (including virtual pseudo-energy) vs. relative loss.

- The optimality "price" one pays for scenario reduction from the original measure P to a simpler measure Q – which can be thought of as adding information that's minimally relevant to the problem in question without actually finding it – can be estimated by the amount $2K\hat{\mu}_{\hat{c}}(P,Q)$. This implies, in particular that one could do a scenario reduction before starting the search for the efficient frontier. In fact, scenario reduction and additional information acquisition are complementary to each other in the sense of information: scenario reduction, as already mentioned can be thought of as an addition of information that's minimally relevant as opposed for information acquisition optimization, where one looks for small amounts of maximally relevant information.

7 Methods for Determining the Efficient Frontier

As was discussed earlier, an approximate solution the optimal information acquisition problem (20) can be greatly facilitated by finding the efficient frontier $L^*(\gamma)$ of maps from Ω into X. The problem of finding the efficient frontier (17) appears to still be sufficiently complex to warrant a search for approximate solutions. Using the result (18) proved in [33], one can reduce the scope of search to subset-optimal maps $g \in \mathcal{C}$ only.

On the other hand, obviously, not all maps in the set $\mathcal{C}$ belong to the set $\mathcal{O}$ of Pareto-optimal map defining the efficient frontier. We call partition $\mathbf{C}$ *optimal* if the corresponding map $g = (\{C_1,\ldots,C_r\},\{x^*_{P_1},\ldots,x^*_{P_r}\}) \in \mathcal{C}$ belongs to the set $\mathcal{O}$ of Pareto-optimal maps. So the problem of finding maps in the set $\mathcal{O}$ is equivalent to that of searching for optimal partitions of the set Ω.

Proposition 4 provides a useful tool for approximating the efficient frontier. We can use the following algorithm (here and later we assume that the original measure P on Ω has a support at a discrete set $\Omega_N \subset \Omega$ consisting of N points).

1. Choose an integer parameter ≥ 2.
2. Choose an appropriate cost function c: $\Omega \times \Omega \to \mathbb{R}_+$ such that $f(\omega,x) \in \mathcal{F}_c$ for all $x \in X$.

Let $\hat{c}$ be the corresponding reduced cost function.

3. Reduce the original measure P to measure Q supported at r points in the set Ω_N, i.e. find a $\hat{c}$-optimal map $\nu \in \mathcal{R}_r(\Omega_N)$ such that $Q = \nu(P)$.

4. Let $\mathbf{C}$ be any partition of Ω generated by the map ν.

5. Let the map $g_{\mathbf{C},P} \in \mathcal{C}$ be a subset-optimal map corresponding to partition $\mathbf{C}$.

Varying the value of parameter r from 2 upwards one can obtain a series of maps in the set $\mathcal{C}$ that are (approximately) Pareto-optimal. Step 2 of the above algorithm is essential for its feasibility. For example, if the problem (1) is a linear multi-period stochastic optimization problem, the cost function of the form (40) can be used. In step 3, finding the measure Q supported at r points that minimizes the value of Monge-Kantorovich functional $\hat{\mu}_{\hat{c}}(P, Q)$ is an NP-hard problem [15] but approximate algorithm such as *fast forward selection algorithm* are available (see Appendix B).

Using the algorithm described above, one can obtain one approximately Pareto-optimal map for each value of the chosen integer parameter. If more Pareto-optimal maps are needed (especially in the region with lower values of pseudo-energy) additional heuristics can be used. For instance, one could begin with the algorithm described above for some relatively high value of r and then merge some of the resulting subsets into one giving rise to a partition with a lower value of r. Clearly, this can be done in $B_r - 1$ ways, where B_n is the n-th Bell number which is just the number of all different partitions of a set consisting of n elements and that can be found from the recursive relation $B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k$ and $B_0 = 1$. (For example, the Bell number for the lower values of n are $B_2 = 2$, $B_3 = 5$, $B_4 = 15$, $B_5 = 52$, $B_6 = 203$, $B_7 = 877$, $B_8 = 4140$.)

We see that if the original chosen value of r is not very high this would lead to a manageable number of partitions. Additionally, scenario reduction can be used to reduce computational complexity of finding the values $x^*_{P_C}$ for subsets C of resulting partitions. On the other hand, if the original value of r makes evaluation of all maps that can be obtained this way computationally prohibitive, a heuristic algorithm described by the following pseudo-code can be used. It finds another partition, with a lower value of r, so that the subset merging procedure can be applied.

The goal of the algorithm represented by the pseudo-code is to identify subsets which are locally compact but as far away from one another as possible. In each step k, we find the average distance $\bar{c}_p$ of each subset center remaining in the index set $J^{[k-1]}$ to only the other remaining centers. The center, and therefore the associated subset, with the largest average distance is chosen and removed from the set $J^{[k-1]}$. The remaining subsets are then merged into a single set.

So far the pseudo-temperature function u has not been taken into account. It is clear, on the other hand, that it will in general affect the composition of the set $\mathcal{O}$ of Pareto-optimal maps. In order to properly incorporate the pseudo-temperature function into the heuristics described above, one could note that the questions difficultly is generally smaller when subsets with high pseudo-temperature values have large measures as well. In other words, if one wishes to keep the question difficulty low, one should avoid creating subsets of small measure in regions of the parameter space characterized with high pseudo-temperature values. To facilitate creation of such subsets, one could, for example modify the (reduced) cost function $\hat{c}$ in the following way

$$\hat{c}(\omega_i, \omega_j) \to \frac{\hat{c}(\omega_i, \omega_j)}{f_c(u(\omega_i), u(\omega_j))}, \tag{27}$$

Algorithm 1: Approximation to Pareto-optimal boundary.

Input:;
$\mathbf{C} = \{C_1, \ldots, C_r\}$,;
$\{\omega_1, \ldots, \omega_r | \omega_i \in C_i\} \subset \Omega$,;
choose an integer n such that $1 \leq n \leq r-2$.;
Step 0**:**;
$J^{[0]} := \{1, \ldots r\}$,;
$\mathbf{C}' := \{C'_1, \ldots, C'_{n+1}\}$ such that $C'_i := \emptyset, \forall i$,;
calculate $\hat{c}_p(\omega_i, \omega_j), \ \forall i, j \in J^{[0]}$.;
Step $k = 1, \ldots, n$**:**;
foreach $i \in J^{[k-1]}$ **do**
$\bar{c}_p(i) := \frac{1}{|J^{[k-1]}|} \sum_{j \in J^{[k-1]}} \hat{c}_p(\omega_i, \omega_j)$,;
end
$u_k := \arg\max_{i \in J^{[k-1]}} \bar{c}_p(i)$,;
$J^{[k]} := J^{[k-1]} \backslash \{u_k\}$,;
$C'_k := C_{u_k}$.;

where f_c: $\mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is some increasing function of its arguments. The specific shape of f_c can be determined experimentally, and several shapes can be tried for every given instance assuming computational resources are not a limiting factor.

8 Examples

Let us consider an example. The original problem is a that of two-stage linear stochastic optimization with simple recourse taken from a well-known texbook [2]. The problem is for a farmer to allocate the appropriate amount of land between wheat, corn and sugar beets in order to maximize profits. The farmer knows that at least 200 tons of wheat and 240 tons of corn must be grown for cattle feed. If not enough is grown to satisfy this demand, both wheat and corn can be bought for \$238 and \$210 per ton, respectively. Any excess above the demand can be sold for \$170 and \$150 per ton of wheat and corn, respectively. It costs \$150 per acre to plant the wheat and \$230 per acre to plant the corn. The farmer can also grow sugar beets that sell for \$36 per ton. However, there is a quota of 6000 tons and any amount grown above this may only be sold at \$10 per ton. It costs \$260 per acre to plant sugar beets. The farmer has 500 acres available.

The problem can be stated as:

$$\begin{aligned} \text{minimize} \quad & 150x_1 + 230x_2 + 260x_3 + \mathbb{E}_P Q(x, \Omega) \\ \text{subject to} \quad & x_1 + x_2 + x_3 \leq 500 \\ & x_1, x_2, x_3 \geq 0, \end{aligned} \tag{FP}$$

where the second stage problem for a specific scenario can be written

$$\begin{aligned} Q(x,s) = \text{minimize}\{&238y_1 - 170w_1 + 210y_2 - 150w_2 - 36w_3 - 10w_4\} \\ \text{subject to} \quad & \omega_1(s)x_1 + y_1 + w_1 \quad \geq 200 \\ & \omega_2(s)x_2 + y_2 + w_2 \quad \geq 240 \\ & w_3 + w_4 \quad \leq \omega_3(s)x_3 \\ & w_3 \quad \leq 6000 \\ & y_1, y_2, w_1, w_2, w_3, w_4 \geq 0 \end{aligned}$$

where $\omega_i(s)$ represents the yield of crop $i := 1, 2, 3$ for wheat, corn, and sugar beets, respectively, under scenario s; x_i are the acres of land to devote to each crop i; y_1, y_2, are tons of wheat and corn, respectively, purchased to meet cattle feed requirements; w_1, w_2, w_3, w_4 are tons of wheat, corn, sugar beets below quota, and sugar beets above quota, respectively, sold for profit.

The problem has been modified in order to create the illustrative example used below. In this example, only wheat and sugar beet yields are uncertain. Each is allowed to take five different values of yields resulting in 25 scenarios. For the sake of convenience, we assume that the corn yield is non-random and is equal to 3 tons per acre, while for both wheat and beets the average yield equal to 2.5 and 20, respectively, has a probability of 0.30. The yield for both of these cultures can be either higher or lower than average by 20% with probability 0.20 and also can be higher or lower than average by 30% with probability 0.15. The yields for wheat and beets are assumed to be independent.

The resulting uncertain yields are summarized below:

wheat (ω_1)	[1.75, 2.00, 2.50, 3.00, 3.25]	w.p. (0.15,0.20,0.30,0.20,0.15)
corn	[3]	w.p. (1)
sugar beets (ω_2)	[14, 16, 20, 24, 26]	w.p. (0.15,0.20,0.30,0.20,0.15)

Also, let us assume that the pseudo-temperature function $u(\omega_1, \omega_2)$ is given as

$$u(i,j) = i \cdot j^{0.5}, \forall \quad i,j \in 1,\ldots,5 \tag{28}$$

where i, j are the indices referencing the uncertain yields of wheat and sugar beets, respectively (where the smallest value of the uncertain yield corresponds to $i = 1$ ($j = 1$) and the largest yield corresponds to $i = 5$ ($j = 5$)). The pseudo-temperature function is then normalized so that $\mathbb{E}_P u(ij) = 1$. Fig. 2 shows a plot of the pseudo-temperature function.

The efficient frontier can be approximated by using the scenario reduction based algorithm described in the previous section together with subset merging heuristics. The resulting maps are shown in Fig. 3 for the case of constant pseudo-temperature. The resulting approximate efficient frontier both for constant pseudo-temperature function and for the pseudo-temperature given shown in Fig. 2 are shown in Fig. 4.

Now consider an information source described by the modified linear model with parameters $b = 0.8$ and $Y_s = 0.2$ (which is a rather modest capacity value). We would like to find out how much the original loss can be reduced by optimally using such an information source. In other words, we want to solve problem (20). For this purpose one can take questions on the (approximate) efficient frontier and plot parametric curves $(Y(\Omega, \mathbf{C}, P, V_\alpha(\mathbf{C})), \mathcal{L}(V_\alpha(\mathbf{C})))$ where $\mathcal{L}(V_\alpha(\mathbf{C}))$ is given by Proposition 2. The question yielding the lowest point of intersection of such a curve with the vertical line $G = Y_s$ will give an approximate solution of problem (20).

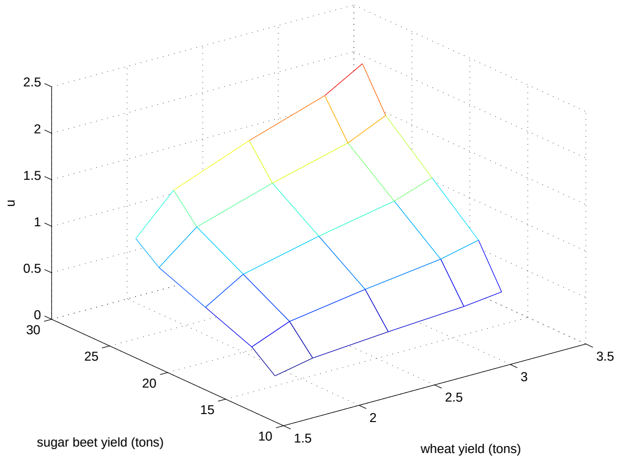

Figure 2: Pseudo-temperature function given for the farmer land allocation problem with uncertainty residing in the yields of wheat and sugar beets.

Results for the case of constant pseudo-temperature are shown in Fig. 5. The parametric curves for three questions (all three with $r = 2$) are produced. We can see that the lowest value of the expected loss that can be obtained this way is equal to 7250 which constitutes a reduction of about 14%.

For the case of non-constant pseudo-temperature are shown in Fig. 6. Analogously, three $r = 2$ questions were chosen on the approximate efficient frontier and the corresponding parametric curves plotted. The best curve is observed to intersect the vertical line $G = 0.2$ at the value of vertical coordinate equal to about 6900 which represents a reduction of about 18% compared to the EVPI of 8450 of the original problem.

9 Conclusion

This paper develops (approximate) methods for solving the problem of optimizing additional information acquisition in decision making problems with uncertainty that are typically solved using stochastic optimization techniques. It represents a logical continuation of the developments presented in [33]. The main problem that was formulated [33] is that of finding an efficient frontier in (pseudo-energy – loss) coordinate plane and to determining the question(s) that would allow to minimize the expected loss for the given (stochastic optimization) problem and a given information source.

The solution methods proposed in this paper are based on the method of probability metrics and their application for scenario reduction in stochastic optimization (see appendices). The main idea is that, informally speaking, optimal scenario reduction on one hand and optimal information acquisition on the other hand are complementary. More specifically, in scenario reduction the goal is reproduce the overall shape of the original probability distribution as faithfully as possible with a small fraction of the original scenarios. In information acquisition, the goal is to identify the types of uncertainty encoded by the original probability distribution the reduction of which would have

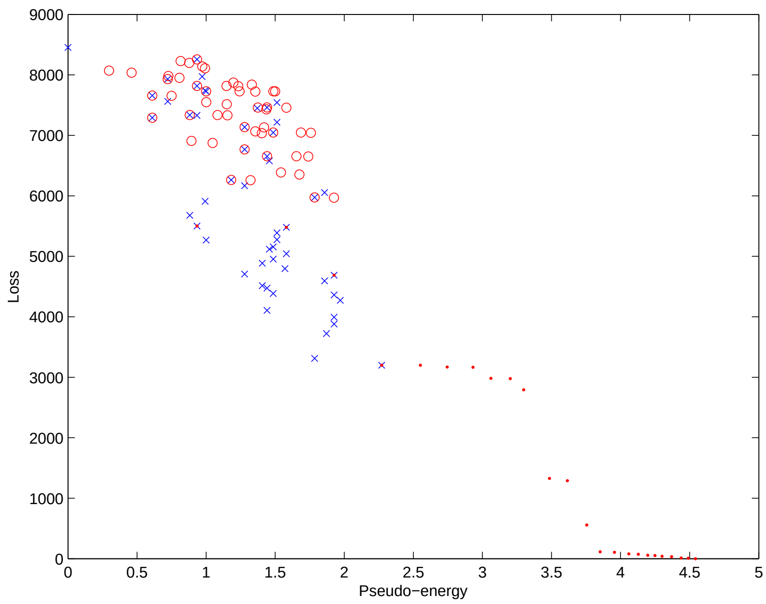

Figure 3: Maps that are generated by scenario reduction for various values of r (solid dots), scenario reduction for $r = 5$ with subsequent subset merging (crosses), scenario reduction to $r = 10$, reducing to $r = 5$ using the pseudo-code and subsequent subset merging (circles). Pseudo-temperature function is set to a constant.

the largest effect on the solution quality. It turns out that these types of uncertainty are associated with the "overall shape" of the distribution (as opposed to "local details") which scenario reduction strives to preserve.

This allows us to develop simple approximate algorithms for determining the efficient frontier (and for finding optimal questions for the given information source) with the help of existing scenario reduction algorithms. The methods described in this paper are shown to work for the class of linear multi-period two stage stochastic optimization problems and should generalize relatively easily to other problem classes for which scenario reduction based on probability metrics was shown to be possible such as chance constrained and two-stage integer stochastic optimization problems.

Appendices

A Probability metrics and stability in stochastic optimization

Consider the problem (1). Let $\mathcal{P}(\Omega)$ be the set of all Borel probability measures on Ω and define

$$v(P) = \inf \left\{ \int_\Omega f(\omega, x)\, dP(\omega) \;:\; x \in X \right\}$$

and

$$S(P) = \left\{ x \in X \;:\; \int_\Omega f(\omega, x)\, dP(\omega) = v(P) \right\}$$

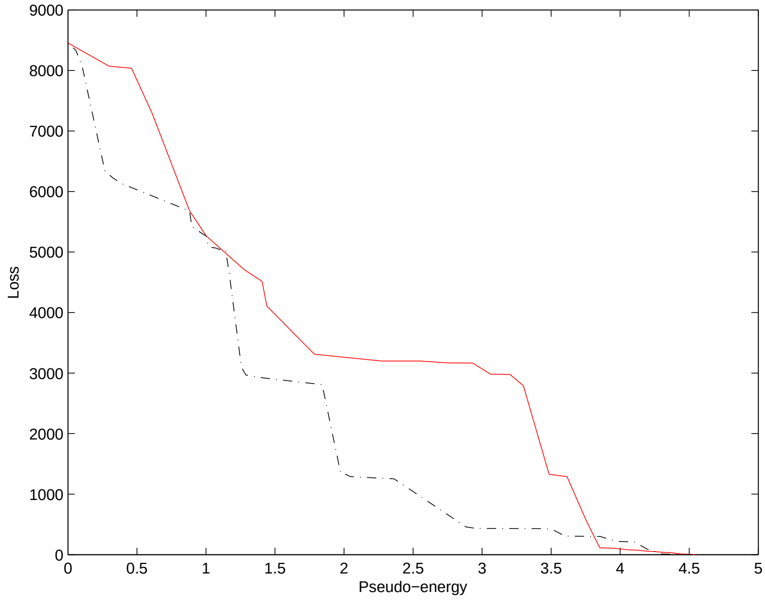

Figure 4: Approximate efficient frontiers for the constant pseudo-temperature function (solid line) and pseudo-temperature shown in Fig. 2.

to be the optimal value and optimal solution set of (1), respectively.

Let's also define (as in, for example, [39])

$$\mathcal{F} = \{f(\cdot, x) : \ x \in X\}$$

and

$$\mathcal{P}_{\mathcal{F}}(\Omega) = \Big\{ Q \in \mathcal{P} : \ -\infty < \int_\Omega \inf_{x \in X \cap \rho\mathbb{B}} f(\omega, x) Q(d\omega) \ \text{ and } \sup_{x \in X \cap \rho\mathbb{B}} \int_\Omega f(\omega, x) Q(d\omega) < \infty, \text{ for all } \rho > 0 \Big\},$$

where $\mathbb{B}$ is the closed unit ball in $\mathbb{R}^n$.

Then the probability distance of the form

$$d_{\mathcal{F},\rho}(P, Q) = \sup_{x \in X \cap \rho\mathbb{B}} \left| \int_\Omega f(\omega, x) P(d\omega) - \int_\Omega f(\omega, x) Q(d\omega) \right| \tag{29}$$

can be defined on $\mathcal{P}_{\mathcal{F}}(\Omega)$. This distance is called *Zolotarev's pseudometric with ζ-structure* [44, 35, 36, 37]. The pseudometric (29) would become a metric if the class $\mathcal{F}$ were rich enough so that $d_{\mathcal{F},\rho}(P, Q) = 0$ implies $P = Q$.

Theorem 2 in [8] states that if $P, Q \in \mathcal{P}_{\mathcal{F}}$, $S(P)$ is nonempty and bounded then there exist $\rho > 0$ and $\delta > 0$ such that

$$|v(P) - v(Q)| \le d_{\mathcal{F},\rho}(P, Q) \tag{30}$$

is valid for all $Q \in \mathcal{P}_{\mathcal{F}}$ such that $d_{\mathcal{F},\rho}(P, Q) < \delta$.

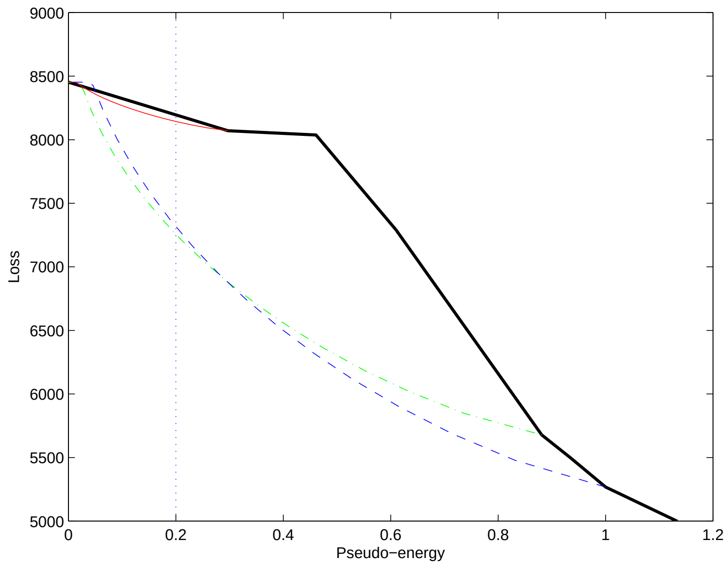

Figure 5: Part of approximate efficient frontier and parametric loss curves for quasi-perfect answers to three different questions for the case of constant pseudo-temperature.

The distance $d_{\mathcal{F},\rho}$ in (30) is typically difficult to handle since the class of functions $\mathcal{F}$ is determined by the specific integrand $f(\omega, x)$ for the given instance of problem (1). The main idea underlying the use of the probability metrics method for the study of stability and for scenario reduction in stochastic programming is to suitably enlarge the class $\mathcal{F}$ so that it still shares its main analytical properties with functions $f(\cdot, x)$. Such properly enlarged classes are sometimes referred to as *canonical classes* and the corresponding metrics are sometimes called *canonical metrics*.

Consider, for instance the class $\mathcal{F}_c$ of continuous functions defined as

$$\mathcal{F}_c = \{f: \ \Omega \to \mathbb{R}: \ |f(\omega) - f(\tilde{\omega})| \le c(\omega, \tilde{\omega}), \text{ for all } \omega, \tilde{\omega} \in \Omega\}, \tag{31}$$

where c: $\Omega \times \Omega \to \mathbb{R}_+$ is a continuous symmetric function such that $c(\omega, \tilde{\omega}) = 0$ if and only if $\omega = \tilde{\omega}$. Then the corresponding (pseudo-) metric has the form

$$\zeta_c(P, Q) \equiv d_{\mathcal{F}_c}(P, Q) = \sup_{f \in \mathcal{F}_c} \left| \int_\Omega f(\omega) P(d\omega) - \int_\Omega f(\omega) Q(d\omega) \right| \tag{32}$$

and is known as *Fortet-Mourier metric*. If the cost function $c(\omega, \tilde{\omega})$ satisfies additional boundedness and continuity conditions:

- $c(\omega, \tilde{\omega}) \le \lambda(\omega) + \lambda(\tilde{\omega})$ for some λ: $\Omega \to \mathbb{R}_+$ mapping bounded sets into bounded sets,
- $\sup\{c(\omega, \tilde{\omega}) \ : \ \omega, \tilde{\omega} \in \mathbb{B}_\epsilon(\omega_0), ||\omega, -\tilde{\omega}|| \le \delta\} \to 0$ as $\delta \to 0$ for each $\omega_0 \in \Omega$, where $\mathbb{B}_\epsilon(\omega_0)$ is the ϵ-ball centered at ω_0,

the Fortet-Mourier metric (32) admits a dual representation as the *Kantorovich-Rubinstein functional* [38]:

$$\zeta_c(P, Q) = \mathring{\mu}_c(P, Q) = \inf \left\{ \int_{\Omega \times \Omega} c(\omega, \tilde{\omega}) \eta(d\omega, d\tilde{\omega}) : \eta \in \mathcal{P}(\Omega \times \Omega), \ \pi_1 \eta - \pi_2 \eta = P - Q \right\}, \tag{33}$$

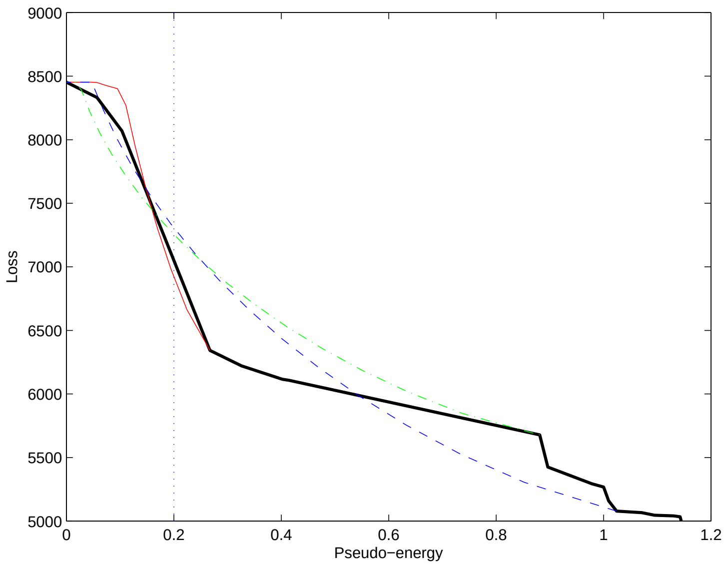

Figure 6: Part of approximate efficient frontier and parametric loss curves for quasi-perfect answers to three different questions for the case of non-constant pseudo-temperature shown in Fig. 2.

where π_1 and π_2 denote projections on first and second components, respectively. It is straightforward to show that the Kantorovich-Rubinstein functional (33) can be upper-bounded by the *Monge-Kantorovich functional*:

$$\mathring{\mu}_c(P,Q) \leq \hat{\mu}_c(P,Q) = \inf\left\{\int_{\Omega\times\Omega} c(\omega,\tilde{\omega})\eta(d\omega,d\tilde{\omega}) : \eta \in \mathcal{P}(\Omega\times\Omega),\ \pi_1\eta = P, \pi_2\eta = Q\right\}, \tag{34}$$

and that the bounds becomes tight, (i.e. $\mathring{\mu}_c(P,Q) = \hat{\mu}_c(P,Q)$) if the cost function $c(\omega,\tilde{\omega})$ is a metric on Ω [26**?**]. The problem of finding the minimum in (34) is known the *Monge-Kantorovich mass transportation problem.*

Note that if measures P and Q are discrete ($P = \sum_{i=1}^N p_i\delta_{\omega_i}$ and $Q = \sum_{j=1}^M q_i\delta_{\tilde{\omega}_j}$), the Monge-Kantorovich functional (34) takes the following form:

$$\begin{aligned}\hat{\mu}_c(P,Q) &= \min\left\{\sum_{i=1}^N\sum_{j=1}^M c(\omega_i,\tilde{\omega}_j)\eta_{ij} : \eta_{ij}\geq 0, \sum_{i=1}^N \eta_{ij} = q_j, \sum_{j=1}^M \eta_{ij} = p_i\ \forall i,j\right\}\\ &= \max\left\{\sum_{i=1}^N p_iu_i + \sum_{j=1}^M q_jv_j : u_i + v_j \leq c(\omega_i,\tilde{\omega}_j)\ \forall i,j\right\}\end{aligned} \tag{35}$$

Given the cost function $c(\omega,\tilde{\omega})$ one can define the *reduced cost* $\hat{c}(\omega,\tilde{\omega})$ on $\Omega\times\Omega$ by

$$\hat{c}(\omega,\tilde{\omega}) = \inf\left\{\sum_{i=1}^{m-1} c(\omega_i,\omega_{i+1}) : m\in\mathbb{N},\ \omega_i\in\Omega,\ \omega_1=\omega,\ \omega_m=\tilde{\omega}\right\}. \tag{36}$$

In can easily be shown that the reduced cost function $\hat{c}(\omega,\tilde{\omega})$ is a metric (since it satisfies the triangle inequality) on Ω and that $\hat{c}(\omega,\tilde{\omega}) \leq c(\omega,\tilde{\omega})$ with the inequality being tight when $c(\omega,\tilde{\omega})$ is also a metric.

It can also be shown (see [36], chapter 4) that if Ω is compact with analytic sublevel sets then the Kantorovich-Rubinstein functional (33) with the reduced cost function $\hat{c}$ coincides with the Kantorovich-Rubinstein functional with the original cost function c (the result referred to as the *reduction theorem*):

$$\mathring{\mu}_{\hat{c}}(P,Q) = \mathring{\mu}_c(P,Q). \tag{37}$$

Since the reduced cost is a metric on Ω we have $\mathring{\mu}_{\hat{c}}(P,Q) = \hat{\mu}_{\hat{c}}(P,Q)$ and, comparing with (37) we conclude that, for compact parameter spaces with analytic sublevel sets, the equality

$$\mathring{\mu}_c(P,Q) = \hat{\mu}_{\hat{c}}(P,Q) \leq \hat{\mu}_c(P,Q) \tag{38}$$

holds true.

We thus arrive at the following useful stability result. If the integrand in problem (**??**) belongs to class $\mathcal{F}_c$ for all $x \in X$ for some cost function c satisfying additional boundedness and continuity conditions described earlier in the appendix, then the estimate

$$|v(P) - v(Q)| \leq \zeta_c(P,Q) = \mathring{\mu}_c(P,Q) = \hat{\mu}_{\hat{c}}(P,Q) \tag{39}$$

is valid for Borel measures P and Q in $\mathcal{P}_c(\Omega)$ on compact Ω characterized with analytic sublevel sets. (Here $\mathcal{P}_c(\Omega) = \{Q \in \mathcal{P}(\Omega) : \int_\Omega c(\omega,\omega_0)dQ(\omega) < \infty\}$ for some $\omega_0 \in \Omega$.)

The particular function $c(\omega,\tilde{\omega})$ that plays an important role in the context of convex stochastic optimization has the form

$$c_p(\omega,\tilde{\omega}) = \max\{1, ||\omega - \omega_0||^{p-1}, ||\tilde{\omega} - \omega_0||^{p-1}\}||\omega - \tilde{\omega}||, \tag{40}$$

for some $\omega_0 \in \Omega$. The corresponding metric $\zeta_p \equiv \zeta_{c_p}$ is referred to as the *p-th order Fortet-Mourier metric.*

To give an example of a class of problems for which the p-th order Fortet-Mourier metric is relevant, consider linear multi-period stochastic optimization problems of the form

$$\min\left\{cy_0 + \mathbb{E}_P\Big(\min\sum_{j=1}^{l} c_j(\omega)y_j\Big) : y_0 \in X,\ y_j \in Y_j,\ W_{jj}y_j = b_j(\omega) - W_{jj-1}(\omega)y_{j-1},\ j = 1,\dots,l\right\}, \tag{41}$$

where $Y_j \subseteq \mathbb{R}^{n_j}$ are polyhedral sets. Problem (41) can be written in the form (1) with the integrand $f(\omega,x)$ given by

$$\begin{aligned} f(\omega,x) &= cx + \inf\left\{\sum_{j=1}^{l} c_j(\omega)y_j : y_j \in Y_j,\ W_{jj}y_j = b_j(\omega) - W_{jj-1}(\omega)y_{j-1},\ j = 1,\dots,l\right\} \\ &= cx + \Psi_1(\omega,x), \end{aligned}$$

where the function $\Psi_1(\omega,x)$ is defined recursively:

$$\Phi_j(\omega,u_{j-1}) = \inf\{c_j(\omega)y_j + \Psi_{j+1}(\omega,y_j) : y_j \in Y_j,\ W_{jj}y_j = u_{j-1}\}$$

$$\Psi_j(\omega,y_{j-1}) = \Phi_j(\omega, b_j(\omega) - W_{jj-1}(\omega)y_{j-1})$$

for $j = l,\dots,1$ and $\Psi_{l+1}(\omega,y_l) \equiv 0$.

It is shown in [39] that if $b_j(\omega) - W_{jj-1}(\omega)x \in W_{jj}Y_j$ for all pairs (ω, x) (*relatively complete recourse*) and $\ker(W_{jj}) \cap Y_j^\infty = \{0\}$ for $j = 1, \ldots, l-1$ (where Y_j^∞ denotes the horizon cone[2] of Y_j) then there exists a constant $\hat{K}$ such that

$$|f(\omega, x) - f(\tilde{\omega}, x)| \leq \hat{K} \max\{1, \rho, ||\omega||^l, ||\tilde{\omega}||^l\} ||\omega - \tilde{\omega}|| \tag{42}$$

for all ω, $\tilde{\omega} \in \Omega$ and $x \in X \cap \rho\mathbb{B}$. This implies that $\frac{1}{\hat{K}\max\{1,\rho\}} f(\omega, x) \in \mathcal{F}_{c_p}$ for all ω, $\tilde{\omega} \in \Omega$ and $x \in X \cap \rho\mathbb{B}$.

It is now straightforward to obtain the following result ([39]). Let $v(P)$ be the optimal value of problem (41). Assume that the relatively complete recourse condition for (41) is satisfied and that $\ker(W_{jj}) \cap Y_j^\infty = \{0\}$ for $j = 1, \ldots, l-1$. Then there exists a constant $K > 0$ such that the estimate

$$|v(P) - v(Q)| \leq K\zeta_{l+1}(P, Q) \tag{43}$$

is valid for any P, $Q \in \mathcal{P}_{l+1}(\Omega)$. (Here $\mathcal{P}_{l+1}(\Omega)$ denotes the set of Borel measures on Ω with finite $(l+1)$-th order moments.)

Specifying the general result (39) to the cost function of the form (40) with $p = l + 1$ we can rewrite the estimate (43) for the difference in optimal objective values of problem (41) as

$$|v(P) - v(Q)| \leq K \mathring{\mu}_{l+1}(P, Q) = K \hat{\mu}_{\hat{c}_{l+1}}(P, Q), \tag{44}$$

where $K > 0$ is some constant.

B Scenario reduction algorithms

The goal of scenario reduction algorithms is, given a stochastic optimization problem of the form (1) characterized by a discrete measure $P = \sum_{i=1}^N p_i \delta_{\omega_i}$ find the discrete measure $Q = \sum_{j=1}^M q_i \delta_{\tilde{\omega}_j}$ such that $M < N$ and the difference in the optimal objective values $|v(P) - v(Q)|$ is as small as possible.

If the stochastic optimization problem has the form (41) of a linear multi-period problem then, as discussed earlier in this section, under relatively complete recourse assumption, the upper bound (44) can be shown to hold. This motivates searching for discrete measures Q that minimize the distance $\hat{\mu}_{l+1}(P, Q)$ (or $\mathring{\mu}_{l+1}(P, Q)$).

Thus the optimal scenario reduction problem based on the method of probability metrics can be formulated as follows [8]. Let $J \subset \{1, 2, \ldots, N\}$ and consider the measure $Q = \sum_{j \notin J} q_j \delta_{\omega_j}$ supported at points ω_j, $j \in \{1, 2, \ldots, N\} \setminus J$. The measure Q is said to be *reduced* from P by deleting scenarios ω_j, $j \in J$ and by assigning new probabilities q_j to the remaining scenarios. The optimal reduction concept proposed in [8] seeks the minimum value of the functional

$$D(J; q) = \hat{\mu}_p \left(\sum_{i=1}^N p_i \delta_{\omega_i}, \sum_{j \notin J} q_j \delta_{\omega_j} \right). \tag{45}$$

[2]The horizon cone D^∞ for the convex set $D \subseteq \mathbb{R}^m$ is defined as the set of all elements $x_d \in \mathbb{R}^m$ such that $x + \lambda x_d \in D$ for all $x \in D$ and all $\lambda \in \mathbb{R}_+$. In particular, $D^\infty = \{0\}$ if D is bounded.

It is shown in [8] that, for set J fixed, the optimal weights q are straightforward to find:

$$q_j = p_j + \sum_{i \in J_j} p_i, \quad \text{for each } j \notin J, \tag{46}$$

where $J_j := \{i \in J : j = j(i)\}$ and $j(i) \in \arg\min_{j \notin J} c_p(\omega_i, \omega_j)$ for each $i \in J$. The corresponding minimum of the functional $D(J; q)$ is

$$D_J = \min_q \{D(J; q) : q_j \geq 0, \sum_{j \notin J} q_j = 1\} = \sum_{i \in J} p_i \min_{j \notin J} c_p(\omega_i, \omega_j).$$

On the other hand, the optimal choice of the set J of given cardinality $|J| = k$

$$\min_J \{D_J = \sum_{i \in J} p_i \min_{j \notin J} c_p(\omega_i, \omega_j) : J \subset \{1, 2, \dots, N\}, |J| = k\}$$

is a combinatorial problem, and it is unlikely that efficient solution algorithms for arbitrary value of k are available. However cases $k = 1$ and $k = N - 1$ are easy to solve to optimality and they can be used to formulate heuristic algorithms for other values of k. The *fast forward* scenario reduction algorithms proposed in [15] proceeds as follows.

Fast forward selection algorithm:

Step 1: $c_{ku}^{[1]} := c_p(\omega_k, \omega_u),\ k, u = 1, \dots, N,$

$$z_u^{[1]} := \sum_{\substack{k=1 \\ k \neq u}} p_k c_{ku}^{[1]},\ u = 1, \dots N,$$

$$u_1 \in \arg \min_{u \in \{1, \dots, N\}} z_u^{[1]},\ J^{[1]} := \{1, \dots, N\} \setminus \{u_1\}.$$

Step i: $c_{ku}^{[i]} := \min\{c_{ku}^{[i-1]}, c_{ku_{i-1}}^{[i-1]}\},\ k, u \in J^{[i-1]},$

$$z_u^{[i]} := \sum_{k \in J^{[i-1]} \setminus \{u\}} p_k c_{ku}^{[i]},\ u \in J^{[i-1]},$$

$$u_i \in \arg \min_{u \in J^{[i-1]}} z_u^{[i]},\ J^{[i]} := J^{[i-1]} \setminus \{u_i\}.$$

Step $n + 1$: Redistribution by (46).

References

[1] Bassamboo, A. and A. Zeevi 2009. On a data-driven method for staffing large call centers. *OR* **57**(3), 514–526.

[2] Birge, J. R. and F. Louveaux 1997. *Introduction to Stochastic Programming*. Springer-Verlag, New York, NY.

[3] Chávez, M., J. Martinerie, and M. Le Van Quyen 2003. Statistical assessment of nonlinear causality: Application to epileptic EEG signals. *J. of Neurosci. Methods* **124**(2), 113–128.

[4] Clemen, R. 1987. Combining overlapping information. *Management Sci.* **33**(3), 373–380.

[5] Clemen, R. and R. Winkler 1999. Combining probability distributions from experts in risk analysis. *Risk Anal.* **19**(2), 187–203.

[6] Cox, R. T. 1946. Probability, frequency, and reasonable expectation. *Am. J. Phys.* **14**, 1–13.

[7] — 1961. *The Algebra of Probable Inference*. Johns Hopkins Press, Baltimore.

[8] Dupačová, J., N. Gröwe-Kuska, and W. Römisch 2003. Scenario reduction in stochastic programming: An approach using probability metrics. *Math. Programming, Ser.A* **95**(3), 493–511.

[9] Ellison, G. and D. Fudenberg 1993. Rules of thumb for social learning. *J. Political Econom.* **101**(4), 612–643.

[10] Fischer, A. J., A. J. Arnold, and M. Gibbs 1996. Information and the speed of innovation adoption. *Amer. J. Agr. Econ.* **78**(4), 1073–1081.

[11] Fisher, M. L. and A. Raman 1996. Reducing the cost of demand uncertainty through accurate response to early sales. *Oper. Res.* **44**(1), 87–99.

[12] French, S. 1985. Group consensus probability distributions: A critical survey. *Bayesian Statist.* **2**, 183–202.

[13] Genest, C. and J. V. Zidek 1986. Combining probability distributions: A critique and an annotated bibliography. *Statist. Sci.* **1**, 114–148.

[14] Harmanec, D. and G. J. Klir 1994. Measuring total uncertainty in Dempster-Shafer theory: A novel approach. *Int. J. Gen. Syst.* **22**(4), 405–419.

[15] Heitsch, H. and W. Römisch 2003. Scenario reduction algorithms in stochastic programming. *Computational Optimization and Applications* **24**(2), 187–206.

[16] Howard, R. A. 1966. Information value theory. *IEEE Transactions on systems science and cybernetics* **2**(1), 22–26.

[17] Huh, W. T. and P. Rusmevichientong 2009. A nonparametric asymptotic analysis of inventory planning with censored demand. *MOOR* **34**(1), 103–123.

[18] Jaynes, E. T. 1957. Information theory and statistical mechanics I. *Phys. Rev.* **106**, 620–630.

[19] — 1957. Information theory and statistical mechanics II. *Phys. Rev.* **108**, 171–190.

[20] Jensen, R. 1988. Information cost and innovation adoption policies. *Management Sci.* **34**(2), 230–239.

[21] Katura, T., N. Tanaka, A. Obata, H. Sato, and A. Maki 2006. Quantitative evaluation of interrelations between spontaneous low-frequency oscillations in cerebral hemodynamics and systemic cardiovascular dynamics. *NeuroImage* **31**(4), 1592–1600.

[22] Klir, G. J. and D. Harmanec 1996. Generalized information theory: Recent developments. *Kybernetes* **25**(7/8), 50–66.

[23] Knuth, K. H. 2005. Lattice duality: The origin of probability and entropy. *Neurocomputing* **67**, 245–274.

[24] Kornish, L. J. and R. L. Keeney 2008. Repeated commit-or-defer decisions with a deadline: The influenza vaccine composition. *Oper. Res.* **56**(3), 527–541.

[25] Levi, R., R. O. Roundy, and D. B. Shmoys 2007. Provably near-optimal sampling-based policies for stochastic inventory control models. *MOOR* **32**(4), 821–839.

[26] Levin, V. L. 1975. On the problem of mass transfer. *Sov. Math. Dokl.* **16**, 1349–1353.

[27] Maeda, Y. and H. Ichihashi 1993. An uncertainty measure with monotonicity under the random set inclusion. *Int. J. Gen. Syst.* **21**(4), 379–392.

[28] McCardle, K. F. 1985. Information acquisition and the adoption of new technology. *Management Sci.* **31**(11), 1372–1389.

[29] Mokhov, I. I. and D. A. Smirnov 2006. El Niño-Southern Oscillation drives North Atlantic Oscillation as revealed with nonlinear techniques from climatic indices. *Geophys. Res. Lett.* **33**. L03708.

[30] Perevalov, E. and D. Grace 2011. Information acquisition for decision making: Answer depth. Submitted for publication.

[31] — 2011. Information acquisition for decision making: Question difficulty. Submitted for publication.

[32] — 2011. Information acquisition for decision making: Source models. Submitted for publication.

[33] — 2011. Optimization of additional information acquisition in decision making problems: main framework. Submitted for publication.

[34] Predd, J. B., D. N. Osherson, S. R. Kulkarni, and H. V. Poor 2008. Aggregating probabilistic forecasts from incoherent and abstaining experts. *Decision Anal.* **5**(4), 177–189.

[35] Rachev, S. T. and W. Römisch 2002. Quantitative stability in stochastic programming: The method of probability metrics. *Math. Oper. Res.* **27**(4), 792–818.

[36] Rachev, S. T. and L. Rüschendorf 1998. *Mass Transportation Problems, Volume I: Theory*. Springer-Verlag, New York.

[37] Rachev, S. T. and L. Rüschendorf 1998. *Mass Transportation Problems, Volume II: Applications*. Springer-Verlag, New York.

[38] Rachev, S. T. and R. M. Shortt 1990. Duality theorems for kantorovich-rubinstein and wasserstein functionals. *Dissertationes Mathematicae* **299**, 647–676.

[39] Römisch, W. and R. J.-B. Wets 2007. Stability of ε-approximate solutions to convex stochastic programs. *SIOPT* **18**(3), 961–979.

[40] Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ.

[41] Shannon, C. E. 1948. A mathematical theory of communication. *Bell Systems Technical Journal* **27**, 379–423, 623–659.

[42] Verdes, P. F. 2005. Assessing causality from multivariate time series. *Phys. Rev. E* **72**. 026222.

[43] Viola, P. A. 1995. Alignment by maximization of mutual information. A.I. Technical Report 1548, Massachusetts Institute of Technology.

[44] Zolotarev, V. M. 1983. Probability metrics. *Theory of Probability and its Applications* **28**(2), 264–287.