Minimizing a stochastic convex function subject to stochastic constraints and some applications

Roiy Jacobovic* Offer Kella*†
June 14, 2019

Abstract

In the simplest case, we obtain a general solution to a problem of minimizing an integral of a nondecreasing right continuous stochastic process from zero to some nonnegative random variable \( \tau \), under the constraints that for some nonnegative random variable \( T, \tau \in [0,T] \) almost surely and \( E\tau = \alpha \) (or \( E\tau \leq \alpha \)) for some \( \alpha \). The nondecreasing process and \( T \) are allowed to be dependent. In fact a more general setup involving \( \sigma \) finite measure, rather than just probability measures is considered and some consequences for families of stochastic processes are given as special cases. Various applications are provided.


AMS Subject Classification (MSC2010): 90C15, 60G99.

1 Introduction

This work presents an approach to solve a certain kind of stochastic programming problems. For general reviews about stochastic programming problems see, e.g., [22, 21]. Given a general probability space, we were initially motivated by finding an optimal random variable \( \tau \) that minimizes

*Department of Statistics and data science; The Hebrew University of Jerusalem; Jerusalem 9190501; Israel. royi.jacobovic@mail.huji.ac.il, offer.kella@gmail.com
†Supported by grant No. 1647/17 from the Israel Science Foundation and the Vigevani Chair in Statistics.
\( E \int_0^T \xi(s)ds \), where \( \xi(\cdot) \) is a nondecreasing right continuous stochastic process such that \( \xi(t) < \infty \) for every \( t \geq 0 \), subject to two types of constraints. The first is \( P(0 \leq \tau \leq T) = 1 \) where \( T \) is some random variable (possibly infinite) which is not necessarily independent of \( \xi(\cdot) \). The second constraint is \( E\tau = \alpha \) where \( \alpha \in [0, ET] \). It turns out that there is a precise and quite clean representation of the optimal \( \tau \) in terms of the pseudo-inverse process associated with \( \xi(\cdot) \). In particular, \( \varphi(t) := \int_0^t \xi(s)ds \) may be non-differentiable with positive probability. This makes the current work related to non-differentiable convex optimization. For references regarding deterministic non-differentiable optimization see, e.g., [5, 18] and Section 11 of [26]. For works about stochastic non-differentiable convex optimization see, e.g., [1, 3, 20]. Section 2 includes the main results of this paper. In fact, they are shown for a somewhat more general setup involving \( \sigma \)-finite measures rather than just probability measures. As to be shown later, this description is useful in various applications to which the other sections are devoted. Section 3 is about the case where \( \xi(\cdot) \) is a deterministic function. Section 4 refers to a special case when \( \xi(\cdot) \) is a strictly increasing linear function with random coefficients. We would like to mention that this problem was our original motivation and an application is given by [7]. Section 5 is about the special case where \( \xi(\cdot) \) does not depend on \( t \). We argue that the results for this case can be applied in hypothesis testing. Section 6 applies the main results in order to solve a static problem which is motivated by the martingale method for solving a dynamic portfolio optimization problem in a continuous time complete market (for details see, e.g., Section 3 of [15]). Section 7 is an application of the current method to find an optimal clearing time for a general clearing model with fixed setup cost and nondecreasing holding cost function (not necessarily linear). For some background on clearing processes see (among others) [11, 12, 13, 23, 24, 25]. In Section 8 we consider the case where \( \xi(\cdot) \) is a renewal counting process and \( T \) is independent of \( \xi(\cdot) \). When \( T \) has an exponential distribution there is a particularly explicit formula for the solution. Finally, Section 9 shows that the current results can be applied to a deterministic setup of separable convex objective function with linear constraints.

2 The main results

Denote \( x^+ = x \vee 0 \), \( x^- = -x \wedge 0 \), where \( x \vee y = \max(x, y) \), \( x \wedge y = \min(x, y) \). Also, for some function \( f \), whenever the limits exist, we denote \( f(t^+) = \lim_{s \uparrow t} f(s) \) and \( f(t^-) = \lim_{s \downarrow t} f(s) \). As usual, \( \mu \)-a.s. abbreviates almost...
surely with respect to some (sigma finite or probability) measure \( \mu \) and for a sigma finite measure space \((X, \mathcal{X}, \mu)\) and \( \mathcal{X} \)-measurable \( \zeta : X \to [-\infty, \infty] \) we denote \( \mu \zeta = \int_X \zeta d\mu \) (whenever either \( \mu \zeta^+ < \infty \) or \( \mu \zeta^- < \infty \)). For the special case where \( \mu \) is a probability measure then we write \( E\zeta \) (expected value) instead of \( \mu \zeta \). Finally, \( \nu \ll \mu \) is for \( \nu \) is absolutely continuous with respect to \( \mu \).

From here on when we write \( \inf \{ t | t \in A \} \) we mean \( \inf \{ t | t \in A \cap [0, \infty) \} \). When \( A \cap (0, \infty) \) is empty, the infimum is defined to be \( \infty \). The following is the main idea that leads to our main result.

**Lemma 1** Let \( \varphi : [0, \infty) \to \mathbb{R} \) be convex, right continuous at zero, with right derivative \( \xi \) (necessarily nondecreasing and right continuous). For \( \lambda \in \mathbb{R} \), denote

\[
\eta(\lambda) = \inf \{ t | \xi(t) \geq \lambda \},
\]

where \( \eta(\lambda) = \infty \) if \( \{ t | \xi(t) \geq \lambda \} \) is empty. For a given \( T \in [0, \infty] \) (possibly infinite) and \( \lambda \in \mathbb{R} \) let

\[
\tau_\lambda = \eta(\lambda) \land T.
\]

If \( \tau_\lambda < \infty \), then for every finite \( t \in [0, T] \)

\[
\varphi(t) \geq \varphi(\tau_\lambda) + \lambda(t - \tau_\lambda).
\]

Moreover, if \( \tau_u < \infty \) for some \( u > \lambda \) then also

\[
\varphi(t) \geq \varphi(\tau_{\lambda+}) + \lambda(t - \tau_{\lambda+}),
\]

where \( \tau_{\lambda+} = \eta(\lambda+) \land T \).

Observe that if we would replace \( \lambda(t - \tau_\lambda) \) on the right side of (3) by \( \xi(\tau_\lambda)(t - \tau_\lambda) \) then the resulting inequality would be an immediate consequence of convexity (since \( \xi(\tau_\lambda) \) is a subgradient at \( \tau_\lambda \)) and would be valid for any choice of \( \tau_\lambda \), not necessarily the one defined in (2). However, for what follows, it is important to have \( \lambda \) rather than \( \xi(\tau_\lambda) \) as the multiplier.

**Proof:** It is well known that \( \xi(t) \geq \lambda \) if and only if \( t \geq \eta(\lambda) \). Therefore, \( \varphi(t) - \lambda t = \int_0^t (\xi(s) - \lambda) ds \) is decreasing on \([0, \eta(\lambda))\) (empty when \( \eta(\lambda) = 0 \)) and nondecreasing on \([\eta(\lambda), \infty)\) (empty when \( \eta(\lambda) = \infty \)). Hence, it is decreasing on \([0, \tau_\lambda)\) and nondecreasing on \([\tau_\lambda, T] \cap [\tau_\lambda, \infty)\). This implies that when \( \tau_\lambda < \infty \), it minimizes \( \varphi(t) - \lambda t \) on \([0, T]\). Thus, for every finite \( t \in [0, T] \) and every \( \lambda \) such that \( \tau_\lambda < \infty \) we have that

\[
\varphi(\tau_\lambda) - \lambda \tau_\lambda \leq \varphi(t) - \lambda t
\]
which is equivalent to (3). Clearly, (4) follows from the continuity of $\varphi$. ■

We now abuse the notation and instead of a function $\xi$, a nonnegative constant $T$ and a constant $\tau_\lambda$, from here on, these would now become functions of the form $\xi(t) = \xi(x,t)$, where we suppress the (functional) dependence on $x$. Although our main concern is with probability spaces (in which case $\xi$ is a stochastic process), it will prove useful to state the following more general result from which everything else follows. This is the main result of this paper. We will abbreviate $\zeta \in \mathcal{X}$ to mean that $\zeta$ is $\mathcal{X}$-measurable and $\tau \in [0,T] \mu$-a.s. to mean that $\tau(x) \in [0,T(x)]$ for $\mu$-almost all $x \in X$.

**Theorem 1** Given a measurable space $(X, \mathcal{X})$ and sigma finite measures $\nu, \mu$ such that $\nu \ll \mu$, assume that $\xi(t) \in \mathcal{X}$ for each $t \geq 0$, $\xi(t) = \xi(x,t)$ is right continuous and nondecreasing in $t$ for each $x \in X$ and $\mu$-a.s. finite for each $t > 0$. Let $T = T(x) \in \mathcal{X}$ be $\mu$-a.s. nonnegative (possibly infinite) satisfying

$$\nu \int_0^T \xi(s)^- ds < \infty .$$

(6)

For $\alpha \in (0,\mu T)$ and $\tau = \tau(x)$, consider

$$\min \nu \int_0^\tau \xi(s) ds$$

s.t. $\tau \in \mathcal{X}$

$$\tau \in [0,T] \mu$-a.s.

$$\mu \tau = \alpha .$$

(7)

Let $Y = d\nu/d\mu$ be a nonnegative finite version of the Radon-Nykodim derivative. With $\tau_\lambda = \inf\{t|Y\xi(t) \geq \lambda\} \wedge T$, if there exists a $\lambda$ satisfying $\mu \tau_\lambda = \alpha$ or $\mu \tau_{\lambda+} = \alpha$, then, respectively, $\tau_\lambda$ or $\tau_{\lambda+}$ solves (7). Otherwise, either $\mu \tau_\lambda = \infty$ for all $\lambda \in \mathbb{R}$ or there exists a $\lambda$ for which $\mu \tau_\lambda < \alpha < \mu \tau_{\lambda+}$. If $\mu \tau_{\lambda+} < \infty$, let

$$q \equiv \frac{\alpha - \mu \tau_\lambda}{\mu \tau_{\lambda+} - \mu \tau_\lambda} .$$

(8)

Then, $(1-q)\tau_\lambda + q\tau_{\lambda+}$ solves (7).

We note that if one prefers that $\xi(t)$ is right continuous and nondecreasing $\mu$-a.s. rather than for every $x \in X$, then in addition one needs to assume in Theorem 1 that $(X, \mathcal{X}, \mu)$ is complete. This is a technical nuisance which we prefer to avoid here.
Proof: We first observe that the assumption $\nu \int_0^T \xi(s)\,ds < \infty$ is needed in order to insure that $\nu \int_0^T \xi(s)\,ds$ is well defined (possibly infinite) for each $\mathcal{X}$-measurable $\tau \in [0, T]$ $\mu$-a.s.

Next we note that since $\xi(t)$ is nondecreasing and right continuous for each $x \in X$ and is $\mu$-a.s. finite for each $t > 0$, then so is $Y\xi(t)$. Thus there is no loss of generality in assuming that $\nu = \mu$ (with $Y = 1$).

We recall that from right continuity it follows that as a function of $(x, t)$, $\xi \in \mathcal{X} \otimes \mathcal{B}[0, \infty]$ (jointly measurable, e.g., Remark 1.4 on p. 5 of [10]). Here $\mathcal{B}$ is for Borel. Thus $\int_0^t \xi(s)\,ds \in \mathcal{X}$ for each $t \geq 0$. Since $\{\tau \leq t\} = \{Y\xi(t) \geq \lambda\} \cup \{T \leq t\}$,

\begin{equation}
\text{this implies that } \tau_\lambda \in \mathcal{X} \text{ for each } \lambda \text{ (and, in fact, that it is jointly measurable as a function of } x, \lambda, \text{ but this will not be needed here).}
\end{equation}

If for some $\lambda$ either $\mu\tau_\lambda = \alpha$ or $\mu\tau_{\lambda+} = \alpha$ then we simply apply one of the inequalities (3), (4) with $t = \tau$ and integrate with respect to $\mu$, observing that $\mu [\lambda(\tau - \tau_\lambda)] = \lambda(\alpha - \alpha) = 0$ or $\mu [\lambda(\tau - \tau_{\lambda+})] = \lambda(\alpha - \alpha) = 0$.

Recall that $\xi(t)$ is $\mu$-a.s. finite for every $t > 0$. From (9) it follows that $\mu$-a.s. $\lim_{\lambda \to -\infty} \tau_\lambda = 0$ and $\lim_{\lambda \to \infty} \tau_\lambda = T$. Thus, from monotone convergence $\mu\tau_\lambda \to \mu T$ as $\lambda \to \infty$ and, when $\mu\tau_\lambda < \infty$ for some $\lambda \in \mathbb{R}$, it converges by dominated convergence to 0 as $\lambda \to -\infty$. Thus, when $\mu\tau_\lambda$ is not infinite for all $\lambda$, for each $\alpha \in (0, \mu T)$ we can take $\lambda$ such that

\begin{equation}
\mu\tau_\lambda < \alpha < \mu\tau_{\lambda+}.
\end{equation}

Assuming that the right side is finite (equivalent to $\mu\tau_u < \infty$ for some $u > \lambda$), then clearly, $(1 - q)\tau_\lambda + q\tau_{\lambda+} \in [0, T]$ $\mu$-a.s. and $\mu [(1 - q)\tau_\lambda + q\tau_{\lambda+}] = \alpha$.

From (3), (4) and the convexity of $\int_0^T \xi(s)\,ds$ in $t$, we have that

\begin{align*}
\mu \int_0^{(1-q)\tau_\lambda + q\tau_{\lambda+}} \xi(s)\,ds &\leq (1 - q)\mu \int_0^{\tau_\lambda} \xi(s)\,ds + q\mu \int_{\tau_\lambda}^{\tau_{\lambda+}} \xi(s)\,ds \\
&\leq (1 - q) \left( \mu \int_0^{\tau} \xi(s)\,ds - \lambda(\alpha - \mu\tau_\lambda) \right) + q \left( \mu \int_0^{\tau} \xi(s)\,ds - \lambda(\alpha - \mu\tau_{\lambda+}) \right) \\
&= \mu \int_0^{\tau} \xi(s)\,ds.
\end{align*}

for every $\tau$ satisfying the constraints, so the proof is complete.

We observe that when $\alpha = 0$, every $\mathcal{X}$-measurable $\tau \in [0, T]$ $\mu$-a.s. with $\mu\tau = 0$ necessarily satisfies that $\tau = 0$ $\mu$-a.s. Similarly, when $\alpha = \mu T < \infty$,
every \(\mathcal{X}\)-measurable \(\tau \in [0, T]\) \(\mu\)-a.s. with \(\mu \tau = \mu T\) necessarily satisfies that \(\tau = T\) \(\mu\)-a.s. Also note that for \(\alpha \notin [0, ET]\) the problem is infeasible. Thus, these cases are trivial.

We note that when \(\xi(t, x)\) is strictly increasing in \(t\) for every \(x\), then \(\tau_\lambda\) is continuous in \(\lambda\). Hence, if \(\mu\tau_\lambda < \infty\) for all \(\lambda\) (e.g., when \(\mu T < \infty\)), then \(\mu\tau_\lambda\) is continuous in \(\lambda\) and for each \(\alpha \in (0, \mu T)\) there is a \(\lambda \in \mathbb{R}\) for which \(\mu\tau_\lambda = \alpha\). Therefore, in this case there is no need to take a convex combination of \(\tau_\lambda\) and \(\tau_{\lambda+}\).

**Proposition 1** When, in addition to the assumptions of Theorem 1, \(\mu\tau_\lambda < \infty\) for all \(\lambda \in \mathbb{R}\) (e.g. when \(\mu T < \infty\)), with \(\tau(\alpha)\) denoting the optimum of (7) (clearly nondecreasing in \(\alpha\)), \(f(\alpha) = \nu \int_0^{\tau(\alpha)} \xi(s)ds\) is a convex function of \(\alpha\) on \([0, \mu T] \cap [0, \infty)\). Moreover, \(\lim_{\alpha \to \mu T} f(\alpha) = \nu \int_0^T \xi(s)ds\) (including the case that \(T\) is not \(\mu\)-a.s. finite) and if \(\nu \int_0^T \xi(s)^+ds < \infty\) for some \(\lambda\) then \(\lim_{\alpha \to T} f(\alpha) = 0\).

**Proof:** If we take \(u \in (0, 1)\) and some finite \(\alpha, \beta \in [0, ET]\), then

\[
\mu((1-u)\tau(\alpha) + u\tau(\beta)) = (1-u)\alpha + u\beta
\]

so that by minimality of \(\tau((1-u)\alpha + u\beta)\) and convexity of \(\int_0^T \xi(s)ds\) in \(t\) we have that

\[
\nu \int_0^{\tau((1-u)\alpha + u\beta)} \xi(s)ds \leq \nu \int_0^{(1-u)\tau(\alpha) + u\tau(\beta)} \xi(s)ds \leq (1-u)\nu \int_0^{\tau(\alpha)} \xi(s)ds + u\nu \int_0^{\tau(\beta)} \xi(s)ds.
\]

Now, we recall (see the proof of Theorem 1) that \(\lim_{\lambda \to -\infty} \tau_\lambda = 0\) and \(\lim_{\lambda \to \infty} \tau_\lambda = T\). This implies both \(\int_0^T \xi(s)ds\) converges to zero as \(\lambda \to -\infty\) and to \(\int_0^T \xi(s)ds\) as \(\lambda \to \infty\). Dominated (for \(\lambda \to -\infty\) and monotone (for \(\lambda \to \infty\)) convergence (separately for \(\xi(s)^+\) and \(\xi(s)^-\)) implies that this also holds for the integral with respect to \(\nu\). Recall that we assume that \(\nu \int_0^T \xi(s)^-ds < \infty\). Also note that since \(\tau(\alpha)\) is nondecreasing in \(\alpha\), then \(\nu \int_0^{\tau(\alpha)} \xi(s)^-ds\) are nondecreasing in \(\alpha\). Thus they have a limit as \(\alpha\) converges to zero or to \(\mu T\) (which for the latter, with \(\xi(s)^+\), could be infinite). Thus if we take \(\alpha(\lambda) = \mu\tau_\lambda\) then the same limits are obtained when \(\lambda \to \pm\infty\).

We observe that if instead the constraint \(\tau \in [0, T]\) a.s. we take \(\tau \in [S, T]\) a.s. where \(S \in \mathcal{X}\) satisfies \(\mu S < \infty\) and \(0 \leq S \leq T\) \(\mu\)-a.s., then upon taking
\(\tilde{\xi}(t) = \xi(S + t)\), \(\tilde{T} = T - S\) and \(\tilde{\alpha} = \alpha - \mu S\), we are back to the original setup. Therefore, Theorem 1 gives a solution for this case as well. Note that for this optimization problem we may take \(\xi(\cdot)\) to be indexed by \(\mathbb{R}\) on and there is no need to assume that \(S,T\) are nonnegative.

Finally we also observe that if \(\mu\) and \(\nu\) are equivalent measures, then the problem (7) may be replaced by a problem in which \(\mu = \nu\), but the equality \(\mu \tau = \alpha\) is replaced by \(\mu A \tau = \alpha\) where \(A\) is strictly positive (and finite). This implies the following two corollaries for two special cases. The first is when \(\mu\) is replaced by a product measure associated with a counting measure and a probability measure and the second is where \(\mu\) is replaced by the product of Lebesgue measure and a probability measure. It will be useful to refer to those in the examples that will appear later. The straightforward proofs are omitted.

**Corollary 1** Given a probability space \((\Omega, \mathcal{F}, P)\), assume that \(\{\xi_i(t) | t \geq 0\}\), are nondecreasing right continuous stochastic processes with \(P(|\xi_i(t)| < \infty) = 1\) for all \(t > 0\) and \(i\) in some finite or countable index set. Let \(T_i\) be nonnegative (possibly infinite) random variables satisfying

\[
\sum_i E \int_0^{T_i} \xi_i(s)^{-} ds < \infty
\]  

(14)

and let \(A_i\) be positive and finite random variables. Consider the following stochastic optimization problem for \(\alpha \in (0, \sum_i ET_i)\).

\[
\min \sum_i E \int_0^{\tau_i} \xi_i(s) ds
\]

s.t. \(\tau_i \in \mathcal{F} , \forall i\)

\(\tau_i \in [0, T_i] \ \text{P-a.s.}, \forall i\)

\[
\sum_i EA_i \tau_i = \alpha
\]

Denote \(\tau_{i,\lambda} = \inf \{t | \xi_i(t) \geq A_i \lambda \} \wedge T_i\). If there exists a \(\lambda\) satisfying \(\sum_i EA_i \tau_{i,\lambda} = \alpha\) or \(\sum_i EA_i \tau_{i,\lambda^+} = \alpha\), then, respectively, \(\tau_{i,\lambda}\) or \(\tau_{i,\lambda^+}\) for all \(i\), solve (15). Otherwise, either \(\sum_i EA_i \tau_{i,\lambda} = \infty\) for all \(\lambda \in \mathbb{R}\) or there exists a \(\lambda\) for which \(\sum_i EA_i \tau_{i,\lambda} < \alpha < \sum_i EA_i \tau_{i,\lambda^+}\). If \(\sum_i EA_i \tau_{i,\lambda^+} < \infty\), let

\[
q = \frac{\alpha - \sum_i EA_i \tau_{i,\lambda}}{\sum_i EA_i (\tau_{i,\lambda^+} - \tau_{i,\lambda})}.
\]

(16)

Then, \((1 - q)\tau_{i,\lambda} + q \tau_{i,\lambda^+}\), for \(1 \leq i \leq n\), solves (15).
Corollary 2  Given a probability space \((\Omega, \mathcal{F}, P)\) and denoting the Lebesgue measure by \(m\) and \(ds = m(ds)\), assume that \(\xi(s, t) = \xi(\omega, s, t)\) is a measurable process (as a function of \((\omega, s)\)) for each fixed \(t \geq 0\) and nondecreasing right continuous for each fixed \((\omega, s) \in \Omega \times [0, \infty)\), with \(P(|\xi_t(t)| < \infty) = 1\) for all \(t > 0\) and \(m\)-almost each \(s \in [0, \infty)\). Let \(T_s = T_s(\omega) \in \mathcal{X} \otimes \mathcal{F}\) with \(P(T_s \geq 0) = 1\) for \(m\)-almost all \(x \in \mathcal{X}\) and assume that
\[
\int_0^\infty E \int_0^{T_s} \xi(s, t)^{-} dt \, ds < \infty . \tag{17}
\]
Finally let \(A_s = A_s(\omega)\) be a measurable process and satisfy \(P(0 < A_s < \infty) = 1\) for \(m\)-almost all \(s \in [0, \infty)\). Consider the following optimization problem.

\[
\min \int_0^\infty E \int_0^{T_s} \xi(s, t) dt \, ds \\
\text{s.t. } \tau_s \text{ is a measurable process} \\
P(0 \leq \tau_s \leq T_s) = 1 \text{ for } m \text{-almost all } s \in [0, \infty)
\]

\[
\int_0^\infty E A_s \tau_s ds = \alpha . \tag{18}
\]

Let \(\tau_{s, \lambda} = \inf \{t | \xi(s, t) \geq A_s \lambda \} \wedge T_s\). If there exists \(\lambda \in \mathbb{R}\) such that either
\[
\int_0^\infty E A_s \tau_{s, \lambda} ds = \alpha \text{ or } \int_0^\infty E A_s \tau_{s, \lambda^+} ds = \alpha \text{ then, respectively, } \tau_{s, \lambda} \text{ or } \tau_{s, \lambda^+},
\]
for \(s \in [0, \infty)\), solve (18). Otherwise, either \(\int_0^\infty E A_s \tau_{s, \lambda} ds = \infty\) for all \(\lambda\) or there exists some \(\lambda\) for which \(\int_0^\infty E A_s \tau_{s, \lambda} ds < \alpha < \int_0^\infty E A_s \tau_{s, \lambda^+} ds\). When \(\int_0^\infty E A_s \tau_{s, \lambda^+} ds < \infty\), denote
\[
q = \frac{\alpha - \mu \int_0^\infty E \tau_{s, \lambda} ds}{\int_0^\infty E A_s (\tau_{s, \lambda^+} - \tau_{s, \lambda}) ds} \tag{19}
\]
and then \((1 - q) \tau_{s, \lambda} + q \tau_{s, \lambda^+}\), for \(s \in [0, \infty)\), solve (18).

Remark 1  It should be observed that upon taking \(\tau = (1 - q) \tau_{\lambda} + q \tau_{\lambda^+}\) in (11) it follows that the first inequality in (11) is actually an equality. Therefore we can take a probability space \(([0, 1], \mathcal{B}([0, 1]), m)\) (where \(m\) is Lebesgue measure) and consider the random variable \(I = 1_{[0, q]}(\omega)\). Then take \((1 - I) \tau_{\lambda} + I \tau_{\lambda^+}\) on the space \(([0, 1] \otimes \mathcal{X}, \mathcal{B}([0, 1]) \otimes \mathcal{X}, P \otimes \mu)\) and obtain that
\[
P \otimes \mu \int_0^{(1-I)\tau_{\lambda} + I \tau_{\lambda^+}} \xi(s) \, ds = (1-q) \mu \int_0^{\tau_{\lambda}} \xi(s) ds + q \int_0^{\tau_{\lambda^+}} \xi(s) ds \\
\leq \mu \int_0^{\tau} \xi(s) ds = P \otimes \mu \int_0^{\tau} \xi(s) ds \tag{20}
\]
for every \( \tau \) satisfying the constraints. In this case we may refer to \((1-I)\tau_\lambda + I\tau_{\lambda+}\) as a mixed strategy.

**Remark 2** It is easy to check that if \( \xi \) is indexed by \( \mathbb{Z}_+ \) (instead of \([0,\infty)\)) and \( \tau \) and \( T \) are integer valued, then the results of this section continue to hold without change with the exception that \( \tau_\lambda \) is defined to be \( \inf\{n: \xi(n) \geq \lambda\} \wedge T \) (integer valued) and that instead of \((1-q)\tau_\lambda + q\tau_{\lambda+}\) (which is not necessarily an integer) we need to take a mixed strategy \((1-I)\tau_\lambda + I\tau_{\lambda+}\) as appearing in Remark 1. Similarly, the same is true for \( \tau_{i,\lambda} \) and \( \tau_{x,\lambda} \).

**Remark 3** Observe that if we replace \( \mu \tau = \alpha \) in (7) by \( \mu \tau \leq \alpha \), then from Lemma 1 it follows that, when finite, \( \tau_0 \) and \( \tau_{0+} \) (as defined in Theorem 1), minimize \( \int_0^T Y \xi(s)ds \) on \([0,T] \cap [0,\infty)\) for each \( x \) and therefore it minimizes the integral with respect to \( \mu \). Thus, if \( \mu A\tau_0 \leq \alpha \) then \( \tau_0 \) is an optimal solution for this modified problem. Otherwise, the optimal solution is the one given in Theorem 1. The reason is that \( \lambda \) for which \( E\tau_\lambda \leq \alpha \leq E\tau_{\lambda+} \) is necessarily negative and thus replacing \( \alpha \) by \( \mu \tau \in [0,\alpha] \) in (11) gives on the right side \( \mu \int_0^T \xi(s)ds - \lambda (\mu \tau - \alpha) \). Since \( \lambda < 0 \) it follows that \( -\lambda (\mu \tau - \alpha) \leq 0 \). This can also be deduced from the convexity reported in Proposition 1. Naturally, the same is valid for Corollaries 1 and 2.

We now proceed to some examples.

### 3 Minimizing a deterministic convex function

When \( \xi \) is deterministic we can conclude the following.

**Corollary 3** Assume that \( \psi: [0,\infty) \to \mathbb{R} \) is strictly convex, right continuous at zero (deterministic) and \( T \) is a nonnegative, finite mean random variable with distribution \( F \). Denote

\[
F_e(t) = \frac{1}{ET} \int_0^t (1 - F(s))ds
\]

(stationary remaining lifetime distribution). Then for every \( p \in (0,1) \)

\[
\tau_p = F_e^{-1}(p) \wedge T \tag{21}
\]

minimizes

\[
\min E\psi(\tau)
\]

s.t. \( \tau \in [0,T] \) a.s.

\[
E\tau = pET .
\]
Proof: Since $\psi$ is strictly convex, then its right derivative $\xi$ is strictly increasing and thus $\eta$ is continuous. Thus, there exists $\lambda$ for which $E\eta(\lambda) \wedge T = pET$ and, by Theorem 1 the optimal solution is $\eta(\lambda) \wedge T$. Now, since $\eta(\lambda)$ is a deterministic constant, then

$$pET = E\eta(\lambda) \wedge T = \int_0^{\eta(\lambda)} (1 - F(s))ds = ETF_e(\eta(\lambda)) \quad (22)$$

from which it follows that $\eta(\lambda) = F_e^{-1}(p)$, where we note that the inverse is well defined since $F_e$ is strictly increasing and continuous on $[0, \sup\{t | F(t) < 1\})$.

We note that when $\psi$ is convex but not strictly convex and for some strictly convex function $\varphi$ on $[0, \infty)$ (hence bounded below) we have that $E\varphi(T) < \infty$, then for any $\tau$ satisfying $\tau \in [0, T]$ a.s. and $E\tau = pET$ we have that $E\varphi(\tau) \leq \varphi(0) \vee E\varphi(T) < \infty$ and (since $\psi + \epsilon\varphi$ is strictly convex)

$$E\psi \left( F_e^{-1}(p) \wedge T \right) + \epsilon E\varphi \left( F_e^{-1}(p) \wedge T \right) \leq E\psi(\tau) + \epsilon E\varphi(\tau) \quad . \quad (24)$$

Upon letting $\epsilon \downarrow 0$ it follows that with the added condition that $E\varphi(T) < \infty$ for some strictly convex function on $[0, \infty)$, Corollary 3 is valid for any convex function $\psi$.

One immediate special case is minimizing $\text{Var}(\tau)$ subject to the constraints $\tau \in [0, T]$ almost surely and $E\tau = pET$ for $p \in (0, 1)$. We also note that when $T$ is constant then it is easy to check that $F_e^{-1}(p) = pT$ and thus $\tau_p = pT$, as expected. This, of course, also follows from Jensen’s inequality as $\psi(\tau_p) = \psi(pT) \leq E\psi(\tau)$ for any $\tau$ with $E\tau = pET$. In contrast, we recall that for this case it is well known that the maximum is given by $IT$ where $I \sim \text{Bern}(p)$. To refresh one’s memory, this follows from

$$\psi(\tau) \leq \left( 1 - \frac{T}{\tau} \right) \psi(0) + \frac{T}{\tau} \psi(T) \quad (25)$$

and then taking expected values, noting that $E\tau/T = p$.

Finally, it is interesting to note that the optimal solution in this section does not depend on the choice of the convex function $\psi$. This is not necessarily so when $\psi$ is stochastic. However, recalling that, for any convex $\psi$, an optimal solution to min $\sum_{i=1}^n \psi(x_i)$ subject to the constraints $x_i \geq 0$ and $\sum_{i=1}^n x_i = \alpha$ is $x_i = \alpha/n$ (which also does not depend on $\psi$), then perhaps we should not be too surprised.
4 Optimizing a quadratic function with random coefficients

Consider the problem
\[ \begin{align*}
\min & E \left( A\tau^2 + B\tau + C \right) \\
\text{s.t.} & \ \tau \text{ is a random variable} \\
& \tau \in [0, T] \text{ a.s.} \\
& ED\tau = \alpha
\end{align*} \] (26)

for any a.s. finite random variable \( A, B, C, D, T \) with \( ET < \infty, EB^- < \infty, EC^- < \infty \) and \( P(A > 0) = P(D > 0) = 1 \) (having an arbitrary joint distribution). Then, the assumptions of Corollary 1 are met with \( n = 1, \xi(t) = 2At + B \) for every \( t \geq 0 \) and
\[ \begin{align*}
\tau_\lambda &= \frac{(D\lambda - B)^+}{2A} \land T = \frac{(\lambda - B/D)^+}{2A/D} \land T.
\end{align*} \] (27)

Thus, for this case, if \( ET < \infty \) then, by continuity (and dominated convergence), for every \( \alpha \in (0, EDT) \) there always is a (finite) \( \lambda \) such that \( ED\tau_\lambda = \alpha \). We also recall that for \( \alpha = 0 \) the solution is a.s. zero and for \( \alpha = ET \) it is a.s. \( T \).

This was applied in [7] to solve a related quadratic optimization problem associated with a certain Lévy storage model. We also note that the special case of a uniform finite sample space probability space results in the (deterministic) quadratic program and its solution reported in [6].

5 Optimizing a linear function with random coefficients

When the goal is to solve the following problem
\[ \begin{align*}
\min & E \left( A\tau + B \right) \\
\text{s.t.} & \ \tau \in [0, T] \text{ a.s.} \\
& EC\tau = \alpha
\end{align*} \] (28)

where \( P(A > 0) = P(C > 0) = 1 \) and \( EB^- < \infty \), we simply take \( \xi(t) = A \) for every \( t \geq 0 \). For this case we have that \( \xi(t) = A \) and thus
\[ \tau_\lambda = \inf \{ t | A \geq C\lambda \} \land T = T_{1_{\{A/C \leq \lambda\}}} \] . (29)
As in the quadratic case, $A, B, C, T$ may have an arbitrary joint distribution.

Note that the same holds in the discrete time case, where we recall Remark 2. In particular, if we take $T = 1$ this results in an alternative (but, admittedly, somewhat less elegant) proof of the Neyman-Pearson Lemma or, more generally, uniformly most powerful tests for this setup, where we would like to test the hypotheses

$$\begin{align*}
H_0 : P &= P_0 \\
H_1 : P &= P_1 ,
\end{align*}$$

(30)

where $P_0, P_1$ are absolutely continuous with respect to a common $\sigma$-finite measure, under either of the constrains $P_0 \tau = \alpha$ or $P_0 \tau \leq \alpha$. See [2] for an in-depth treatise of such (and more general) problems which exploits convexity. In particular, compare equations (1.8)-(1.12) there to what appears here.

6 Relation to portfolio selection

In this section we refer to the classical model of dynamic utility maximization with consumption in a complete continuous-time stock market which is presented in Section 3 of [15]. There it is shown that one approach to solving the dynamic problem is to first solve the following static problem

$$\begin{align*}
\text{maximize}_{\{c_t: t \in [0,T]\}, X} &\quad E \left[ \int_0^T U_1(t, c_t) dt + U_2(X) \right] \\
\text{subject to} &\quad 0 \leq X , 0 \leq c_t , \forall t \in [0,T] , \ P \ - \ a.s.
\end{align*}$$

(31)

where $T, x_0 > 0$ are constants, $\{H_t; t \in [0,T]\}$ is a certain $P$-a.s. positive stochastic process whose distribution is determined endogenously by the model setup and defined at the beginning of Section 2.3 of [15]. Importantly, this process is not influenced by the decision variables of the optimization. In addition, $U_1$ and $U_2$ are deterministic functions such that for every $c \geq 0$, $U_1(\cdot, c)$ is continuous and for every $t \in [0,T]$, $U_1(t, \cdot)$ and $U_2(\cdot)$ are assumed to be $C_1$ strictly concave functions that their derivatives satisfy some additional conditions. In addition, we also consider the case where either $U_1$ or $U_2$ is identically zero. With these assumptions, as mentioned by [9], even when $U_1$ is identically zero, Lagrange multipliers for this problem may not exist and hence Lagrange method is not always applicable. Another solution which is based on the exact definition of the process $\{H_t; t \geq 0\}$ provided by
Section 3.4 of [15]. Solving (31) is covered by our Corollary 2. In particular, one does not need to assume any differentiability assumptions on the utility functions, as is usually assumed in this literature. Also, the assumption that the utility functions are deterministic functions can also be relaxed. Namely, take $U_2(\omega,t)$ to be a.s. convex in $t$ for each $\omega$. The same can be done with respect to $U_1$. Now, let $S$ and $V$ be two nonnegative random variables. In addition assume that $\{\Gamma_t; t \in [0,T]\}$ and $\{\Upsilon_t; t \in [0,T]\}$ are two nonnegative stochastic processes. In particular, assume that these random quantities are exogenous to the model, i.e. they are not influenced by the choice of $X$ and $\{c_t; t \geq 0\}$. The requirement that $X$ and $\{c_t; t \geq 0\}$ are nonnegative could be replaced by the constraint

$$X \in [S,V], c_t \in [\Gamma_t, \Upsilon_t], \forall t \in [0,T], \quad P-a.s.$$  

to which the results of Section 2 still apply. Examples of models considering such constraints are, e.g., [8, 14, 16, 17]. Another case which is also covered by the current work is when $T = \infty$ and $U_2$ is identically zero. Finally, note that this kind of optimizations is also motivated by discrete time models (see, e.g., Section 3.3 of [4]).

7 Optimal clearing times in a regenerative clearing process

If $\xi$ is a nonnegative process then we can think of $\{\xi(t)|0 \leq t < \tau\}$ as the first cycle of a (regenerative) clearing process. When $E \tau < \infty$, for such a clearing process an ergodic distribution exists and if $\xi^*$ has this distribution then we have that for any nonnegative Borel $g$,

$$E g(\xi^*) = \frac{1}{E \tau} E \int_0^\tau g(\xi(s)) ds .$$  

(32)

Note that if $g$ is a nonnegative, nondecreasing and right continuous function then $g(\xi(\cdot))$ is a nonnegative, nondecreasing right continuous process and we can apply the results of Section 2 to optimize $E g(\xi^*)$ subject to the constraints in (7). This also provides a method for solving the following optimization problem for any given $K > 0$ and nonnegative, nondecreasing right continuous $g$:

$$\min \frac{1}{E \tau} \left( K + E \int_0^\tau g(\xi(s)) ds \right)$$

s.t. $\tau \in (0,T]$ a.s. .  

(33)
In this case the cost structure is a setup cost $K$ incurred right after each clearing and a (possibly nonlinear) holding cost function $g$. The solution is obtained by first restricting the minimization to feasible $\tau$’s satisfying $E\tau = \alpha \in (0, ET]$ to obtain

$$h(\alpha) = \frac{1}{\alpha} \left( K + E \int_0^{\tau(\alpha)} g(\xi(s))ds \right)$$  \hspace{1cm} (34)$$

where $\tau(\alpha)$ denotes the optimal solution from Theorem 1 and then $h$ is minimized over $(0, ET]$ either analytically, when possible, or numerically.

Finally we observe that for every $a, b \geq 0$ such that $a + b > 0$ and every $x, y$ we have that

$$\frac{ax + by}{a + b} \geq x \land y.$$  \hspace{1cm} (35)$$

Setting

$$a = (1 - q)E\tau_\lambda$$
$$b = qE\tau_{\lambda^+}$$

$$x = \frac{1}{E\tau_\lambda} E \int_0^{\tau_\lambda} \xi(s)ds$$
$$y = \frac{1}{E\tau_{\lambda^+}} E \int_0^{\tau_{\lambda^+}} \xi(s)ds,$$

(see (8)) we infer that

$$\frac{1}{\alpha} E \int_0^{\tau(\alpha)} \xi(s)ds \geq \frac{1}{E\tau_\lambda} E \int_0^{\tau_\lambda} \xi(s)ds \land \frac{1}{E\tau_{\lambda^+}} E \int_0^{\tau_{\lambda^+}} \xi(s)ds \hspace{1cm} (37)$$

implying that

$$\inf_{\tau \in (0, T]} \frac{1}{E\tau} E \int_0^{\tau} \xi(s)ds = \inf_{\alpha \in (0, ET]} \frac{1}{\alpha} E \int_0^{\tau(\alpha)} \xi(s)ds$$
$$= \inf_{\lambda \mid E\tau_\lambda > 0} \frac{1}{E\tau_\lambda} E \int_0^{\tau_\lambda} \xi(s)ds.$$  \hspace{1cm} (38)$$

Note that $E\tau_\lambda > 0$ if and only if $P(\tau_\lambda > 0) > 0$. Since $\eta(\lambda) > 0$ if and only if $\lambda > \xi(0)$, it is easy to check that this is equivalent to

$$P(\xi(0) < \lambda, T > 0) > 0.$$  \hspace{1cm} (39)$$
8 Example: Renewal counting process $\xi$ with independent $T \sim \exp(\mu)$ and a bit more

Assume that $T \sim \exp(\theta)$ is independent of $\{\xi(t); t \geq 0\}$ which is a renewal counting process with inter-renewal times distributed like some $X$. As usual, it is assumed that $P(X \geq 0) = 1$ and $P(X = 0) < 1$. Since $\xi(0) = 0$, then $\eta(\lambda) = 0$ for every $\lambda \leq 0$ and, for every $\lambda \geq 0$,

$$\eta(\lambda) = \inf\{t \geq 0; \xi(t) \geq \lambda\} = \inf\{t \geq 0; \xi(t) \geq \lceil \lambda \rceil\} = S_{\lceil\lambda\rceil}$$

where $S_{\lceil\lambda\rceil}$ is the $\lceil\lambda\rceil$th renewal time. Therefore, for every $\lambda \geq 0$, $\tau_{\lambda} = T \land S_{\lceil\lambda\rceil}$, so that

$$E\tau_{\lambda} = T \land S_{\lceil\lambda\rceil} = E \int_0^{S_{\lceil\lambda\rceil}} e^{-\theta t} dt = \frac{1}{\theta} \left(1 - Ee^{-\theta S_{\lceil\lambda\rceil}}\right)$$

For integer valued $\lambda$ we have that $\tau_{\lambda+} = \tau_{\lambda+1}$ and otherwise $\tau_{\lambda+} = \tau_{\lambda}$. It is easily verified that with

$$\lambda_{\alpha} = \left\lfloor \frac{\log (1 - \theta_\alpha)}{\log Ee^{-\theta X}} \right\rfloor,$$

for $\alpha \in (0, \theta^{-1})$, we either have that $E\tau_{\lambda_{\alpha}} = \alpha$ or $E\tau_{\lambda_{\alpha}+} = E\tau_{\lambda_{\alpha}+1} = \alpha$ or $E\tau_{\lambda_{\alpha}} < \alpha < E\tau_{\lambda_{\alpha}+1}$, in which case the optimal solution is $(1-q)\tau_{\lambda_{\alpha}} + q\tau_{\lambda_{\alpha}+1}$ where $q$ is given by (8).

If $T$ is independent of $\xi$ but does not have an exponential distribution, then

$$E\tau_{\lambda} = \int_0^\infty (1 - F_T(t))(1 - F_X^{\lceil\lambda\rceil}(t)) dt = ET \int_0^\infty f_e(t)(1 - F_X^{\lceil\lambda\rceil}(t)) dt$$

where $f_e(t) = (1 - F_T(t))/ET$ and $F_T$ and $F_X$ are the cumulative distribution functions of $T$ and $X$, respectively. In this case there is no explicit formula for $\lambda_{\alpha}$, but in many cases it can be computed numerically. A case which is worth pointing out is when $X \sim \exp(\theta)$. In this case it can be easily verified that

$$E\tau_{\lambda} = ET \sum_{k=0}^{\lceil\lambda\rceil-1} Ee^{-\theta T_e}(\theta T_e)^k/k!$$

(44)
where $T_e$ has a distribution with density $f_e$. In this case

$$Ee^{-\theta T_e} = \frac{1 - Ee^{-\theta T}}{\theta ET}$$

and the $k$th derivative of this function with respect to $\theta$ is given by

$$(-1)^k E e^{-\theta T_e} T_e^k.$$  

Thus, in principal, the knowledge of $Ee^{-sT}$ for every $s \geq 0$ gives us a procedure for finding everything that is needed in order to compute the optimal $\tau$ in this case.

### 9 Separable convex optimization with linear constraints

Obviously, Corollary 1 can be applied to the following optimization problem in which $f_i$ are convex (not necessarily differentiable) functions, $t_i$ nonnegative reals (possibly infinite) and $a_i$ are strictly positive and finite.

$$\min \sum_{i=1}^{n} f_i(x_i) \quad \text{s.t. } x_i \in [0,t_i] \cap [0,\infty), \forall 1 \leq i \leq n$$

When $f_i$ are differentiable, $t_i = \infty$, $a_i = 1$ and $\alpha = 1$, the results are consistent with the famous Gibb’s Lemma, noting that if $\xi_i(0) \geq \lambda$ then necessarily $\eta_i(\lambda) = 0$. This is a standard convex optimization problem with a separable objective function and linear constraints and the number of references is huge (for the case where $f_i$ are differentiable). For example, quite a few examples are given in [19]. The standard solution (under differentiability assumptions) is by applying the Karush-Kuhn-Tucker conditions or Gibb’s Lemma.

### References


