On singular control problems, the time-stretching method, and the weak-M1 topology

Asaf Cohen*

May 8, 2019

Abstract

We consider a general class of singular control problems with state constraints. Budhiraja and Ross [8] established the existence of optimal controls for a relaxed version of this class of problems by using the so-called ‘time-stretching’ method and the J1-topology. We show that the weak-M1 topology is better suited for establishing existence, since by using it, one bypasses the need for time-transformations, without any additional effort. Furthermore, we reveal how the time-scaling feature in the definition of the weak-M1 distance embeds the time-stretching method’s scheme. This case study suggests that one can benefit from working with the weak-M1 topology in other singular control frameworks, such as queueing control problems under heavy traffic.

Keywords: weak-M1 topology, singular control, time-stretching, optimal controls.

AMS subject classification: 93E20, 60J60, 60F99, 60K25.

1 Introduction

In this paper, we revisit the problem of proving the existence of optimal controls for a class of singular control problems with state (and control) constraints. The paper includes two main theorems. Theorem 2.1 argues that there exist optimal singular controls for the problem. Our proof for this theorem uses weak convergence arguments under Skorokhod’s weak-M1 (WM1) topology. This problem was analyzed by Budhiraja and Ross [8] using the standard Skorokhod’s J1 topology and the time-stretching method, which provides tightness of some time-scaled processes and uses rescaling of the time in the limit. We, on the other hand, show that, under the WM1 topology, tightness can be directly obtained for the original sequence. This leads to a simpler proof. Furthermore, in the second main theorem, Theorem 3.1, we shed light on the relationship between the time-stretching method and the WM1 topology.

1.1 Singular control problems

Singular control problems are control problems where the control $dU(t)$ is allowed to be singular with respect to the Lebesgue measure $dt$. Such problems have been studied in various fields

*Department of Statistics, University of Haifa, Haifa, 31905, Israel. Email: shloshim@gmail.com, web: https://sites.google.com/site/asafcohen/tau/. This research was supported by the ISF (grant No. 1169/18).
such as queueing systems, mathematical finance, actuarial science, manufacturing systems, etc. State constraints for singular controlled diffusion processes are natural in many practical problems. For example in queueing systems, the diffusion scaled queueing problem is often approximated by the so-called Brownian control problems (see, [17, 15]). In this case, buffers cannot be negative and in case they have bounded capacity, further restrictions appear. The singular controls are also restricted, see [19]. In the area of mathematical finance and actuarial sciences the prices of assets are often modeled by diffusion processes and the singular controls are often restricted in some way (e.g., non decreasing processes).

We consider the following multi-dimensional problem. Fix a finite horizon $T > 0$. Let $U$ be a process whose increments belong to a closed cone that is contained in an open half-space (for example the nonnegative orthant, which is the case in optimal dividend payouts [1, Section 3], portfolio selection with transaction costs [11, Section 3], and the reduced Brownian network in [19, Section 5]). Specifically, the state process is given by

$$X(t) = x + \int_0^t b(s, X(s)) ds + \int_0^t \sigma(s, X(s)) dW(s) + \int_{[0,t]} k(s) dU(s), \quad t \in [0, T],$$

such that $X(t)$ belongs to a closed and convex set, where $W$ is a Wiener process. The decision maker aims to minimize a cost that accounts for the state process and the singular control. This is a generalization of the model considered in [8] since the coefficients in the state dynamics that we consider are not constants and the state process is not restricted to live in a closed convex cone. Another minor difference is that we consider a finite horizon problem instead of an infinite horizon discounted one. Nevertheless, this is only due to a personal taste of the author. The techniques in this paper can be transferred to the discounted case without any difficulty.

Optimal solutions in such problems, in dimension one, are often defined using Skorokhod’s reflection mapping, see e.g., [5] and [16]. The latter paper’s approach works in higher dimensions as well. The technique is to study the regularity of relevant solutions of a free boundary differential equation. The smoothness is necessary for verifying that the candidate reflected control is optimal, see e.g., [33, 31]. The difficulty with this approach is that regularity is not always available. Hence, in the general case, the value function is characterized as the unique viscosity solution to the associated differential equation, see [2]. Another approach is the time-stretching method on which we now detail.

### 1.2 The time-stretching method

The time-stretching method was introduced by Meyer and Zheng in [25] and studied in the same framework by Kurtz in [21]. In the context of stochastic control, the method was first used by Kushner and Martins in [24, 22] and was adopted in [7, 8, 9, 10]. We now sketch the basic idea of the proof of the existence of optimal singular controls using the time-stretching method. First, one chooses a sequence of asymptotic optimal controls $\{U^n\}$. In case of absolutely continuous controls with a bounded control set, compactness and tightness arguments yield the existence of optimal control. This is not the case with singular controls under the $J_1$ topology, in which case the oscillation can be very big. However, it is possible to scale the time in such a way so that the time-scaled controls (and other relevant processes) are uniformly Lipschitz. The scaled processes are therefore tight and one may consider a limit point of these processes. By
rescaling back to the original time-scale any of the limit points are shown to be optimal by showing the convergence of the costs passing through the time-scaled processes.

1.3 The advantage of the WM1 topology over the J1 topology

Before introducing the WM1 topology, we state a nice feature that makes it so useful in the context of singular controls: the WM1 oscillation of any nondecreasing (component-wise) function is zero! This is in contrast to the J1 oscillation, which can be very big, especially in the existence of jumps. Therefore, in case that the singular controls have nondecreasing increments in each of its components, the proof of the existence of optimal control is fairly easy and only requires probabilistic growth bounds to attend tightness and convergence of the costs. The assumption that the increments are nondecreasing is not restrictive since the increments take values in a closed cone, strictly contained in an open half space. Such a cone can be linearly transformed into the nonnegative orthant and the problem can be reformulated accordingly.

The simplicity of the proof of the existence of optimal singular control demonstrates the advantage of the WM1 topology for singular control problems in any dimension. This suggests that one can benefit from working with this topology in other singular control problems, such as the general approximation of the Brownian control problem to queueing control problems presented in the seminal work of Budhiraja and Ghosh [7], the integral transformation used by Atar and Shifrin [3 Section 3], the Knightian uncertainty model given in [10], and in other queuing models as well. At this point, one may wonder how come the simplicity of the proof using the WM1 topology does not violate the principle that there is no such thing as a free lunch. The reason is that part of the complexity is embedded in the properties of the WM1 topology, established in [36 Section 12]. Hence, one can think of the WM1 topology framework as a ready-made lunch. On the next subsection we explain how the time-stretching method’s scheme is embedded in the definition of the WM1 topology.

1.4 The WM1 topology and its relationship with the time-stretching method

In his seminal paper [32], Skorokhod introduced four ways to evaluate distances in the space of functions that are right-continuous with left limits (RCLL), known as J1, J2, M1, and M2. The associated topologies are named the same. Later, Whitt [36 Section 12] presented strong and weak versions of the M1 and the M2 topologies, which, in each case, coincide in dimension one. The strong topology agrees with the standard topology introduced by Skorokhod and the weak topology coincides with the product topology. While the J1 topology is the most commonly used, over the years, several works have been done under the M1 topology, see e.g., [34, 37, 20, 23, 28, 29, 30, 35, 27, 12, 14, 26]. These works often consider one-dimensional processes, hence the terms weak and strong topologies coincide. The WM1 topology is much less common, yet is still found to be useful, see e.g., [18, 36, 4].

The strong- and weak-topologies over the time interval [0, T] are defined by the following distance

\[ d(x^1, x^2) := \inf_{(\hat{x}^1, \hat{r}^1), (\hat{x}^2, \hat{r}^2)} \left\{ \sup_{0 \leq s \leq 1} |\hat{x}^1(s) - \hat{x}^2(s)| \vee \sup_{0 \leq s \leq 1} |\hat{r}^1(s) - \hat{r}^2(s)| \right\}, \]
where the infimum is taken over all possible continuous nondecreasing functions \((\hat{x}^i, \hat{r}^i), i = 1, 2,\) satisfying \((\hat{x}^i(0), \hat{r}^i(0)) = (x^i(0), 0)\) and \((\hat{x}^i(1), \hat{r}^i(1)) = (x^i(T), T)\), when \((\hat{x}^i, \hat{r}^i)\) traces out the graph of \((x^i(t), t)\) from \(t = 0\) to \(t = T\). Such a mapping \((\hat{x}^i, \hat{r}^i)\) is called a weak parametric representation of \(x^i\). Loosely speaking, \(\hat{x}^i\) is a time scaled version of \(x^i\) with respect to the time scaling function \(\hat{r}^i\). The difference between the strong- and weak-topologies lies in the way the graph is defined. The exact definition of the graph in the WM1 topology appears in Section 3 below.

In Theorem 3.1 we show that the parametric representations embeds the time-stretching method. We consider a sequence of RCLL functions \(\{x^n\}_n\) that converges in the WM1 topology to an RCLL function \(x\). Then, for every \(n\) we construct a weak parametric representation \((\hat{x}^n, \hat{r}^n)\), using the same structure used in the time-stretching method, where recall that the function \(\hat{r}^n\) is the time-scaling function and \(\hat{x}^n\) is the time-scaled function. These functions are uniformly Lipschitz over \(n\), hence a limit point \((\hat{x}, \hat{r})\) exists. We show that, in some sense, \((\hat{x}, \hat{r})\) is a weak parametric representation of \(x\). More accurately, we show that \(x = \hat{x} \circ r\), where \(r\) is the right-inverse of \(\hat{r}\). That is, \(r\) brings the limit of the time-scaled functions \(\hat{x}\) back to the scale of \(x\). This is the same procedure done in the time-stretching method, only that now it is built in the WM1 topology.

1.5 Preliminaries and notation

We use the following notation. The sets of natural and real numbers are respectively denoted by \(\mathbb{N}\) and \(\mathbb{R}\). For any \(m \in \mathbb{N}\), and \(a, b \in \mathbb{R}^m\), \(a \cdot b\) denotes the dot product between \(a\) and \(b\), and \(|a| = (a \cdot a)^{1/2}\) is the Euclidean norm. For \(a, b \in \mathbb{R}\), set \(a \lor b := \max\{a, b\}\) and \(a \land b := \min\{a, b\}\). The interval \([0, \infty)\) is denoted by \(\mathbb{R}_+\). For any interval \(I \subseteq \mathbb{R}\) and any \(m \in \mathbb{N}\), \(\mathcal{D}_d^I := \mathcal{D}(I, \mathbb{R}^d)\) denotes the space of \(\mathbb{R}^d\) valued functions that are RCLL defined on \(I\). For \(f \in \mathcal{D}_d^I\) and \(t \in I\), \(|f| := \sup_{s \in I \cap (-\infty, t]} |f(s)|\). For any event \(A\), \(1_A\) is the indicator of the event \(A\), that is, \(1_A = 1\) if \(A\) holds and 0 otherwise. We use the convention that the infimum of the empty set is \(\infty\).

1.6 Organization

The rest of the paper is organized as follows. In Section 2 we present the singular control problem, state Theorem 2.1 that deals with the existence of optimal singular control theorem, and explicitly introduce the steps of the time-stretching method given in 3. Section 3 is dedicated to the WM1 topology, where we also and show its relationship with the time-stretching method (Theorem 3.1). Finally, in Section 4 we prove the existence theorem using the WM1 topology.

2 The control problem and the main result

Throughout the paper we fix a finite horizon \(T > 0\) and dimensions \(d, d_1, d_2 \in \mathbb{N}\). Let \(\mathcal{U}\) be a closed and convex cone of \(\mathbb{R}^{d_1}\) and let \(\mathcal{X}\) be a closed and convex subset of \(\mathbb{R}^{d_2}\), both with nonempty interiors. Let \(k : [0, T] \to \mathbb{R}^{d \times d_1}\) be a continuous mapping and denote its image by \(K\). Denote \(K\mathcal{U} = \{ku : k \in K, u \in \mathcal{U}\}\) and impose the following assumption.
Assumption 2.1 There are vectors \( v_1 \in \mathbb{R}^d \) and \( u_1 \in \mathbb{R}^{d_1} \) and a parameter \( a_0 > 0 \) such that, for all \( v \in \mathcal{KU} \) and \( u \in \mathcal{U} \),
\[
 v \cdot v_1 \geq a_0 |v| \quad \text{and} \quad u \cdot u_1 \geq a_0 |u|. \tag{2.1}
\]

The geometric interpretation of this assumption is that both \( \mathcal{U} \) and \( \mathcal{KU} \) are subsets of open half-spaces.

Definition 2.1 An admissible singular control for any \( x \in \mathcal{X} \) is a tuple
\[
 \Xi := (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,T]}, \mathbb{P}, X, W, U),
\]
where \( (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,T]}, \mathbb{P}) \) is a filtered probability space satisfying the usual conditions and supporting the processes \( X, W, \) and \( U \) that satisfy the following conditions.

- \( W \) is a \( d_2 \)-dimensional \( \mathcal{F}_t \)-measurable Wiener process;
- \( U = (U(t))_{t \in [0,T]} \) is an RCLL \( \mathcal{F}_t \)-progressively measurable process whose increments take values in \( \mathcal{U} \);
- \( X = (X(t))_{t \in [0,T]} \) is the state process whose dynamics are given by
\[
 X(t) = x + \int_0^t b(s, X(s))ds + \int_0^t \sigma(s, X(s))dW(s) + \int_{[0,t]} k(s)dU(s), \tag{2.2}
\]
and \( X(t) \in \mathcal{X} \) for every \( t \in [0,T] \), where \( b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d \) and \( \sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d \times d_2} \) are measurable functions satisfying further properties given in Assumption 2.2 below.

Throughout the paper we fix \( x \) and denote by \( \mathcal{A} \) the collection of all admissible singular controls. More than often we abuse notation and refer to \( \mathbb{P} \) as the control, denoting \( \mathbb{P} \in \mathcal{A} \) instead of \( \Xi \in \mathcal{A} \). By convention we assume \( U(0^-) = 0 \) and \( X(0^-) = x \).

The cost function associated with the admissible control \( \mathbb{P} \in \mathcal{A} \), is given by
\[
 J(\mathbb{P}) := \mathbb{E}^\mathbb{P}\left[ \int_0^T f(t, X(t))dt + \int_{[0,T]} h(t)dU(t) + g(X(T)) \right],
\]
where \( f : [0, T] \times \mathbb{R}^d \to \mathbb{R} \), \( h : [0, T] \to \mathbb{R}^{d_1} \), and \( g : \mathbb{R}^d \to \mathbb{R} \) are measurable functions that satisfy further properties, given in Assumption 2.2 below. The associate value is
\[
 V := \inf_{\mathbb{P} \in \mathcal{A}} J(\mathbb{P}).
\]

A control \( \mathbb{P} \) is called an optimal control if its associated cost attains the value, that is, \( J(\mathbb{P}) = V \).

The following assumption is needed for the main result of this section to hold:

Assumption 2.2 1. The functions \( b, \sigma, k, f, h, \) and \( g \) are continuous on their domains. Hence, \( k \) and \( h \) are bounded.

2. There exists a constant \( c_k > 0 \) such that \( |\bar{k}u| \geq c_k |u| \) for all \( (\bar{k}, u) \in \mathcal{K} \times \mathcal{U} \).
3. The functions $b$ and $\sigma$ are uniformly bounded and Lipschitz continuous in $x \in \mathbb{R}^d$, uniformly in $t \in [0, T]$.

4. $\{h \cdot u : h \in \mathcal{H}, u \in \mathcal{U}\} \subset \mathbb{R}_+$, where $\mathcal{H}$ is the image of $h$.

5. At least one of the following conditions holds:
   (a) There exist positive constants $C_g, \bar{C}_g$ and $C_f$ and $\bar{p} > p \geq 1$ such that for any $(t, x) \in [0, T] \times \mathcal{X}$,

   \[-C_g(1 - |x|^p) \leq g(x) \leq \bar{C}_g(1 + |x|^p), \tag{2.3}\]

   and

   \[|f(t, x)| \leq C_f(1 + |x|^p).\]

   (b) There exists $C_h > 0$ such that for all $u \in \mathcal{U}$, $h \cdot u \geq C_h |u|$.

Remark 2.1 Pay attention that the conditions in the present paper are more general than the ones imposed in \cite{5} as we now list.

(i) The condition (2.1) is the same as \cite[(1)]{5}, only that in our case $\mathcal{X}$ is a general closed and convex set and not necessarily a cone.

(ii) The dynamics of the state process $X$ in our case follow a general diffusion process plus a singular control component, where the coefficients are not necessarily constants as in \cite{5}. At this point it is worth mentioning that the Lipschitz continuity of $b$ and $\sigma$ are required for the existence of a solution to (2.2).

Aside these generalizations there is another small difference between the models. We study a finite horizon problem with terminal cost and not a discounted one. Therefore, we impose condition (2.3) on the terminal cost and not on the running cost.

Theorem 2.1 Under Assumptions 2.1 and 2.2 the singular control problem admits optimal controls.

The proof of the theorem is provided in Section 4.

2.1 Solving the problem using the time-stretching scheme

We now shortly review the scheme of the time-stretching method used in \cite{5} to prove the existence of optimal controls. The reason for this introduction is two-fold. First, in order to compare between our proof of existence using the WM1 topology and the proof using the time-stretching scheme we need to introduce the scheme (Section 4); and second, we need the notion of time-stretching in order to tie between it and the WM1 topology. The latter is done in Section 3.2.

The idea of the scheme is to consider a sequence of singular controls $\{U^n\}_n$, whose associated payoff converges to the value function. A limiting control (if exists) is a candidate for an optimal control. The problem is that an arbitrary sequence of singular controls is not necessarily relatively compact under the J1 topology since the J1-oscillation can be very big. To deal with this problem one follows the next five-step scheme, which is also illustrated in Figure 4.
(i) Approximate the controls $U^n$ by continuous ones, which are denoted again by $U^n$. For this, one needs the full power of \([8, (1)]\).

(ii) For each $n \in \mathbb{N}$ define the process

$$\tau^n(s) := s + U^n(s) \cdot u_1. \quad (2.4)$$

It is strictly increasing and continuous by (2.1) and since $U^n$ is continuous. Hence, the left inverse $\hat{\tau}^n(t) = \inf\{s \geq 0 : \tau^n(s) > t\}$ is continuous and strictly increasing.

(iii) Define the time-stretched process

$$\hat{U}^n(t) = U^n(\hat{\tau}^n(t)), \quad (2.5)$$

and similarly set $\hat{X}^n$ and $\hat{W}^n$. The paths of these processes are uniformly Lipschitz, hence $C$-tightness (under the $J_1$ topology) is attained, and one can consider a limit point $(\hat{U}, \hat{X}, \hat{W}, \hat{\tau}) = \lim_{k \to \infty} (\hat{U}^{n_k}, \hat{X}^{n_k}, \hat{W}^{n_k}, \hat{\tau}^{n_k})$, along a converging subsequence indexed by $\{n_k\}_k$.

(iv) Define the time-inverse process $\tau(t) := \inf\{s \geq 0 : \hat{\tau}(s) > t\}$, and set up the rescaled process

$$U(t) = \hat{U}(\tau(t)), \quad t \in [0, T], \quad (2.6)$$

and similarly for $X$ and $W$.

(v) Show that $U$ is an optimal control for the original problem by proving convergence of the costs, passing through the time-stretched and rescaled processes.

\[
\begin{array}{cccc}
U^n(t) & \xrightarrow{\hat{\tau}^n} & \hat{U}^n(t) = U^n(\hat{\tau}^n(t)) & \leftarrow \tau \\
\quad \text{does not converge} & \quad \text{under the $J_1$ topology} & \quad \text{under the $J_1$ topology} & n \to \infty \\
U(t) = \hat{U}(\tau(t)) & \xrightarrow{\tau} & \hat{U}(t) = U(\hat{\tau}(t)) & \\
\end{array}
\]

Figure 1: The scheme of the time-stretching method applied to the continuous control $U^n$.

Pay attention that the approximating continuous controls from the first step imply that $\tau^n$ is continuous in addition to increasing, and hence $\hat{\tau}^n$ is strictly increasing. This property is very convenient when going back to the original scale. Saying this, it is mentioned in \([8]\) (without a proof) that this approximation is not necessary\(^1\).

\(^1\)Cohen \([10]\) avoided the approximation by continuous controls and managed to bypass this issue because in the queueing model considered there the singular control process has small jumps, by nature. Hence in the limit the oscillation of the time-stretched process is small even though it is not continuous.
3 The WM1 topology

We now set up the WM1 topology on $\mathcal{D}_{[0,T]}^m$ and state a few results that serve us in the sequel. For additional reading about this topology and the other Skorokhod topologies, the reader is referred to [36]. Please note that Whitt also sets up the strong M1 topology, which we ignore in this manuscript. The reason is that our framework involves multidimensional processes for which the strong-M1 topology is not useful and in the one-dimensional case the weak- and strong-M1 topologies coincide.

3.1 The weak parametric representation

Fix $m \in \mathbb{N}$. For any $a, b \in \mathbb{R}^m$ define the product segment

$$[[a, b]] := [a_1, b_1] \times \ldots \times [a_m, b_m] \subset \mathbb{R}^m,$$

where $[a_i, b_i] := [a_i \land b_i, a_i \lor b_i] = \{\alpha a_i + (1 - \alpha) b_i : 0 \leq \alpha \leq 1\}$. For any $x \in \mathcal{D}_{[0,T]}^m$ define the thick graph of $x$ by

$$G(x) := \{((z, t) \in \mathbb{R}^m \times [0, T] : z \in [[x(t_1), x(t)]])\}
= \{((z, t) \in \mathbb{R}^m \times [0, T] : z \in [x_i(t_1), x_i(t)], 1 \leq i \leq m\},
$$

where $x(0-)$ is $x(0)$. A weak (partial) order relation is defined on the graph $G(x)$ as follows:

$$(z^1, t^1) \leq (z^2, t^2) \text{ if either } t^1 < t^2 \text{ or } t^1 = t^2 \text{ and for all } i, |x_i(t_1^1) - z_i^1| \leq |x_i(t_1^2) - z_i^2|.$$

The WM1 topology is defined by a semi-metric $d_w$ (does not satisfy the triangle inequality, see [36] Example 12.3.2). To set it up, define the weak parametric representation of $x$ to be a continuous nondecreasing (with respect to the weak order defined above) function $(\hat{x}, \hat{r})$ mapping $[0, 1]$ into $G(x)$ such that $(\hat{x}(0), \hat{r}(0)) = (x(0), 0)$ and $(\hat{x}(1), \hat{r}(1)) = (x(T), T)$. The component $\hat{r}$ scales the time interval $[0, T]$ to $[0, 1]$ and $\hat{x}$ time-scales $x$. Let $\Pi_w(x)$ be the set of all the weak parametric representations of $x$. Define,

$$d_w(x^1, x^2) := \inf_{(\hat{x}, \hat{r}) \in \Pi_w(x^j), j=1,2} \{||\hat{x}^1 - \hat{x}^2||_1 \lor |\hat{r}^1 - \hat{r}^2|_1\}.$$

Pay attention that the parametric representations bring $x^1$ and $x^2$ to the same time-scale $[0, 1]$, hence the parametric representations are ‘comparable’. A nice observation that serves us in the sequel is that if one sets the right-inverse of $\hat{r}$, $r(s) := \inf\{t \geq 0 : \hat{r}(t) > s\} \land 1$, then

$$\hat{x}(r(t)) = x(t), \quad t \in [0, T]. \quad (3.2)$$

The hat notation is consistent with the one given in Section 2.1.

Remark 3.1 (A general construction of a parametric representation, [36] Remark 12.3.3) The basic idea for the parametrization is to ‘stretch’ the time in a way that for every jump of $x$ we associate a subinterval of $[0, 1]$ on which the scaled time component $\hat{r}$ stays constant and $\hat{x}$ increases (with respect to the partial order defined above) to match the values of $x$ at the endpoints of the chosen subinterval. Explicitly, let $\{t_j\}_j \subset [0,T]$ be the set of all the discontinuities of $x$. For each $j$ pick a subinterval $[a_j, b_j] \subset [0,1]$, $a_j < b_j$. For every $s \in [a_j, b_j]$ set $\hat{r}(s) = t_j$ and let $\hat{x} : [a_j, b_j] \to [[x(t_j-), x(t_j)]$ be nondecreasing with respect to the partial
order, such that \( \hat{x}(a_j) = x(t_j^-) \) and \( \hat{x}(b_j) = x(t_j) \), (for example, Whitt suggested to take \( \hat{x} \) to be defined via a linear interpolation between \((a_j, x(t_j^-))\) and \((b_j, x(t_j))\)). Do this in a way that \( t_j < t_k \) holds if and only if \( b_j < a_k \). Let \( t \) be a continuity point of \( x \). If \( t \) is a limit of a subsequence of discontinuity points \( \{t_k\}_k \), set up \( \hat{r}(a) = t \) and \( \hat{x}(a) = \lim_{k \to \infty} x(t_k^-) \), where \( a = \lim_{k \to \infty} a_k \) and \( \hat{r}(a_k) = t_k \). Finally, we are left with a collection of open intervals of the form \((a, b)\) on which \((\hat{x}, \hat{r})\) is not defined. We use linear interpolation and set up

\[
\hat{r}(t) = \frac{b - t}{b - a} r(a) + \frac{t - a}{b - a} r(b), \quad \hat{x}(t) = x(r(t)), \quad t \in (a, b).
\]

Pay attention that whenever a jump occurs, the time is stretched. This is the first hint for the connection we aim to establish in the next section.

### 3.2 The relationship between WM\(1\) and the time-stretching

Before establishing the relationship in the general case, we provide an example for WM\(1\) convergence \(d_w(x^n, x) \to 0\) in \(D^2_{[0,2]}\). This example also clarifies the weak parametric representation. The numbers are taken from \[36, \text{Example 12.3.1}\], where it is also shown that the convergence does not hold under the strong-M1 topology (which we ignore). The weak parametric representations that we choose to work with are different than the ones Whitt used. The reason is that we construct parametric representations in the same way the time is scaled in the time-stretching scheme, given in Section 2.1.

#### 3.2.1 An illuminating example

Let \( x, x^n : [0, 2] \to \mathbb{R}^2 \) be given by

\[
x(s) = \begin{cases} 
(0, 0), & s \in [0, 1), \\
(2, 2), & s \in [1, 2],
\end{cases}
\quad x^n(s) = \begin{cases} 
(0, 0), & s \in [0, 1 - \frac{1}{n}), \\
(2, 1), & s \in [1 - \frac{1}{n}, 1), \\
(2, 2), & s \in [1, 2].
\end{cases}
\]

The first observation is that \( G(x) = G(x^n) = \left[\left(0, 0\right), (2, 2)\right] = [0, 2] \times [0, 2] \). Hence, weak parametric representations only need to satisfy the continuity and monotonicity condition for \((\hat{x}, \hat{r})\) and \((\hat{x}^n, \hat{r}^n)\), with the initial-terminal conditions \((\hat{x}(0), \hat{r}(0)) = (\hat{x}^n(0), \hat{r}^n(0)) = ((0, 0), 0) \) and \((\hat{x}(1), \hat{r}(1)) = (\hat{x}^n(1), \hat{r}^n(1)) = ((2, 2), 2) \). To this end we construct the representation by stretching the time whenever a jump occurs in the same manner done in the time-stretching scheme, given in Section 2.1. The connection between this scheme and the parametric representation is discussed extensively immediately after the example. Following (2.4), define the function \( r^n : [0, 2] \to [0, 1] \) by

\[
r^n(s) = \begin{cases} 
\frac{1}{6} s, & s \in [0, 1 - \frac{1}{n}), \\
\frac{1}{6} (s + 3), & s \in [1 - \frac{1}{n}, 1), \\
\frac{1}{6} (s + 4), & s \in [1, 2].
\end{cases}
\]
Next, define the left inverse of \( r^n \) by

\[
\hat{r}^n(t) = \inf \{ s \geq 0 : r^n(s) > t \} \land 1 = \begin{cases} 
6t, & t \in [0, \frac{1}{6}(1 - \frac{1}{n})], \\
1 - \frac{1}{n}, & t \in [\frac{1}{6}(1 - \frac{1}{n}), \frac{1}{6}(1 - \frac{1}{n}) + \frac{1}{2}], \\
6t - 3, & t \in [\frac{1}{6}(1 - \frac{1}{n}) + \frac{1}{2}, \frac{5}{3}], \\
1, & t \in [\frac{5}{3}, \frac{n}{6}], \\
6t - 4, & t \in [\frac{5}{6}, 1]. 
\end{cases}
\]

Also, set \( \hat{x}^n = (\hat{x}_1^n, \hat{x}_2^n) \) as follows.

\[
\hat{x}^n(t) = \begin{cases} 
(0, 0), & t \in [0, \frac{1}{6}(1 - \frac{1}{n})], \\
(2, 1)(2t - \frac{1}{3}(1 - \frac{1}{n})), & t \in [\frac{1}{6}(1 - \frac{1}{n}), \frac{1}{6}(1 - \frac{1}{n}) + \frac{1}{2}], \\
(2, 1), & t \in [\frac{1}{6}(1 - \frac{1}{n}) + \frac{1}{2}, \frac{5}{3}], \\
(2, 1) + (0, 1)(6t - 4), & t \in [\frac{5}{3}, \frac{5}{6}], \\
(2, 2), & t \in [\frac{5}{6}, 1]. 
\end{cases}
\]

Pay attention that \( \hat{x}_1^n \) and \( \hat{x}_2^n \) increase only when \( \hat{r}^n \) is flat. The structure of \( (\hat{x}^n, \hat{r}^n) \) is consistent with the scheme given in Remark 3.1.

Now, the elements of the sequence \( \{(\hat{x}^n, \hat{r}^n)\}_n \) are uniformly Lipschitz and uniformly converge to \( (\hat{x}, \hat{r}) \), given by,

\[
\hat{x}(t) = \begin{cases} 
(0, 0), & t \in [0, \frac{1}{6}), \\
(2, 1)(2t - \frac{1}{3}), & t \in [\frac{1}{6}, \frac{2}{3}), \\
(2, 1) + (0, 1)(6t - 4), & t \in [\frac{2}{3}, \frac{5}{6}), \\
(2, 2), & t \in [\frac{5}{6}, 1]. 
\end{cases}
\]

\[
\hat{r}(t) = \begin{cases} 
6t, & t \in [0, \frac{1}{6}), \\
1, & t \in [\frac{1}{6}, \frac{5}{6}), \\
6t - 4, & t \in [\frac{5}{6}, 1].
\end{cases}
\]

This is indeed a weak parametric representation of \( x \). Pay attention that it is only pathwise linear along the interval \([\frac{1}{6}, \frac{5}{6})\), on which \( \hat{r} \) is constant, and not linear. This is allowed of course by the definition of the thick graph. As can be seen from Figure 2, the form of \( \hat{x} \) is inherited by the forms of \( \{\hat{x}_n\}_n \).

We now discuss about the connection between the weak parametric representation and the time-stretching scheme, which is given in Section 2.1.

**Remark 3.2** The functions \( x^n, x, r^n, r, \hat{r}, \hat{x}^n, \) and \( \hat{x} \) are equivalent versions of \( U^n, U, \tau^n, \hat{\tau}, \hat{r}, \hat{U}^n \) := \( U^n(\hat{x}^n) \), and \( \hat{U} \), respectively. Indeed, pay attention that for any \( a, b \geq 0, (1, 1) \cdot (a, b) > \| (a, b) \| \), hence \( [2.1] \) holds with \( u_1 = (1, 1) \) and \( u_0 = 1 \). However, there are some differences that follow one after the other. The first one is that we construct a weak parametric representation for the noncontinuous function \( x^n \) itself and not for an approximating continuous function. This yields the second difference: the functions \( \hat{r}^n \) are not strictly

---

\(^2\)This is in fact what distinguishes the weak- from the strong-topology. In the strong topology, the thick graph is replaced by a thin graph (see the definition in [36, 12.3.3]) and our \( \hat{x} \) cannot be a part of a strong parametric representation. The reason is that under the strong topology, it must be linear along \( t \in [1/6, 5/6] \) connecting \( (0, 0) \) and \( (2, 2) \). The gap between this two functions indicate that the convergence holds only under WM1.
Figure 2: The two graphs on top are of $\hat{r}^n$ and $\hat{r}$ and the ones at the bottom are of $\hat{x}^n = (\hat{x}_1^n, \hat{x}_2^n)$ and $\hat{x} = (\hat{x}_1, \hat{x}_2)$, respectively. The two solid red lines in the graphs of $\hat{x}^n$ and $\hat{x}$ describe the mapping from $[0, 1]$ to $\mathbb{R}^2$ and the dashed blue lines represent the marginals. Pay attention that $\hat{x}^n$ (resp., $\hat{x}$) is non-constant only when $\hat{r}^n$ (resp., $\hat{r}$) is constant and that $\hat{x}^n$ is in fact linear when $\hat{r}^n$ is flat, while $\hat{x}$ is only pathwise linear, when $\hat{r}$ is flat. In fact, on the interval $[1/6, 5/6]$, $\hat{x}$ can be seen as a linear interpolation with the additional $t$-value point $4/6$. This point is inherited from the prelimit functions $\hat{x}^n$, $n \in \mathbb{N}$. 
Hence, one can set up \( \hat{x}^n \) with parameterization of \( x \) \( r \) and not by \( x^n \circ \hat{r}^n \) as done in the stochastic problem when setting up \( \tilde{U}^n = U^n \circ \hat{r}^n \). Nevertheless, when going back to the original scale both methods work in the same way. In point (iv) in Section 2.1 one defines \( U(t) = \tilde{U}(\tau(t)) \), where \( \tau(t) = \inf\{ s \geq 0 : \hat{r}(s) > t \} \) and in our setting as well \( x(t) = \hat{x}(r(t)) \), where \( r(t) = \inf\{ s \geq 0 : \hat{r}(s) > t \} \wedge T \), see (3.2). The minimum with \( T \) comes since we consider a finite time-horizon, unlike [3].

This comparison confirms that indeed in the stochastic model one may avoid the continuous controls approximation (Step (i)) in Section 2.1 and define the time-stretched processes using linear interpolation on intervals where \( \hat{r}^n \) is constant. Clearly, the notation becomes heavier in this case and this procedure is less favorable than the one that asserts Step (i).

3.2.2 Establishing the relationship in the general case

The arguments for the general case are similar and are now explicitly provided. Consider a relatively compact sequence \( \{x^n\}_n \subset D^m_{[0,T]} \), which is also uniformly bounded in total variation. The latter requirement is essential for the time-stretching method to hold, hence we assume it holds here as well. By reducing to a subsequence, which is relabeled by \( \{n\} \), consider \( x \) such that \( d\lambda(x^n,x) \to 0 \) as \( n \to \infty \). We now set up a weak representation in the same way done in the last example, which is consistent with the definitions of \( r^n \) and \( \hat{r}^n \) from (2.4). Denote by \( M^n \) the total variations of \( |x^n| \) over \([0,T]\), and set

\[
  r^n(s) := \frac{1}{M^n} \left( t + \int_0^t d|x^n(s)| \right).
\]

Define its left-inverse \( \hat{r}^n(t) := \inf\{ s \geq 0 : r^n(s) > t \} \wedge 1 \). The first observation is that \( \hat{r}^n(0) = 0, \hat{r}^n(1) = T \), and \( \hat{r}^n \) is nondecreasing. Next, pay attention that \( \hat{r}^n \) jumps together with \( x^n \). Each jump of \( x^n \) then leads to a corresponding interval on which \( \hat{r}^n \) is constant. Hence, one can set up \( \hat{x}^n \) as suggested in Remark 3.1 so that \( (\hat{x}^n, \hat{r}^n) \) is a weak-representation parameterization of \( x^n \). By the right continuity of \( r^n \) and \( \hat{r}^n \), the following identity holds

\[
  \hat{x}^n(r^n(t)) = x^n(t).
\]

This is the equivalent of (2.5). Indeed, the continuity of \( U^n \) there implies that \( \hat{r}^n(\tau^n(t)) = \tau^n(\hat{r}^n(t)) = t \), hence \( \tilde{U}^n(\tau^n(t)) = U^n(\hat{r}^n(\tau^n(t))) = U^n(t) \). The functions \( \{(\hat{x}^n, \hat{r}^n)\}_n \) are uniformly Lipschitz, hence this sequence is relatively compact. Consider a limit point \( (\hat{x}, \hat{r}) \) attained along a subsequence, which is relabeled by the indexes \( \{n\} \). In the next theorem we show that in the limit \( n \to \infty \), one obtains the equivalent of (2.6). Hence, establishing the desired connection. This is illustrated in Figure 3. Compare it with Figure 1.

**Theorem 3.1** The functions \( \hat{x}^n \circ r^n \) converge to \( \hat{x} \circ r \) uniformly over \([0,T]\), and moreover, \( \hat{x} \circ r = x \).

**Proof.** Since \( \{\hat{x}^n\}_n \) are uniformly Lipschitz and \( \hat{x} \) is its limit, the convergence \( \hat{x}^n \to \hat{x} \) is uniform over \([0,T]\). Hence,

\[
  \lim_{n \to \infty} \sup_{0 \leq t \leq T} |\hat{x}^n(r^n(t)) - \hat{x}(r^n(t))| = 0.
\]
\[
x^n(t) \xrightarrow{n \to \infty} \hat{x}^n(t) \quad \text{converges under the WM1 topology}
\]

Theorem 3.1: \[\hat{x}(r(t)) = x(t)\]

Figure 3: The scheme of the parametric representations. Theorem 3.1 establishes that \(\hat{x}(r(t)) = x(t)\). Compare it with Figure 1.

According to [13, Theorem 3.5.6], the space \(D^m_{[0,T]}\) is separable. Hence, there is a sequence of step functions \(\{\hat{y}^m\}_m\) that uniformly converges to \(\hat{x}\) as \(m \to \infty\). Hence,

\[
\lim_{m \to \infty} \sup_{0 \leq t \leq T} |\hat{x}(r^n(t)) - \hat{y}^m(r^n(t))| = 0 \quad \text{and} \quad \lim_{m \to \infty} \sup_{0 \leq t \leq T} |\hat{x}(r(t)) - \hat{y}^m(r(t))| = 0.
\]

Together with (3.5), in order to establish the uniform convergence in the theorem, it is sufficient to show that for any fixed \(m\),

\[
\lim_{n \to \infty} \sup_{0 \leq t \leq T} |\hat{y}^m(r^n(t)) - \hat{y}^m(r(t))| = 0.
\]

And the last convergence follows since the uniform convergence \(\hat{r}^n \to \hat{r}\) implies \(\int_0^T 1_{\{r^n(t) \in [a,b]\}} dt \to \int_0^T 1_{\{r(t) \in [a,b]\}} dt\), as \(n \to \infty\), for any \(0 \leq a < b \leq T\). This establishes the uniform convergence in the theorem.

We now show that \(\hat{x}(r(t)) = x(t)\) on \([0, T]\). Recalling that \(\lim_{n \to \infty} d_w(x^n, x) = 0\), it follows by the definition of \(d_w\) (and also explicitly stated in Proposition 3.1 (iii) below) that for every \(t\) which is a continuity point of \(x\), the convergence \(x^n(t) \to x(t)\) holds. From (3.4) and the first part of the theorem, it follows that \(x^n(t) \to \hat{x}(r(t))\). The last two limits imply that \(\hat{x} \circ r = x\) on a dense subset of \([0, T]\). Since both functions are right-continuous, they are identified by their values on a dense set, and the last part of the theorem follows.

\[\square\]

### 3.3 Oscillation and compactness

In this section we set up some oscillation functions and use them in order to establish compactness results, which are necessary for the proof of Theorem 2.1. Some of the results in [36], e.g., Theorem 12.12.2, assert that the functions in \(D^m_{[0,T]}\) are continuous at the boundary points \(t = 0\) and \(t = T\). This assertion is not part of the requirements for \(U\) (thus nor for \(X\) as well) mentioned in Definition 2.1 hence should be avoided. Furthermore, recall that we assumed \(U(0-) = 0\) and \(X(0-) = x\). To this end, we slightly modify some of the definitions given in [36] and work on a closed interval whose interior contains \([0, T]\), for simplicity we consider the interval \([-1, T + 1]\). Set up

\[
\tilde{D}^m_{[0,T]} := \{ x \in D([-1, T + 1], \mathbb{R}^m) : x(t) = 0 \text{ for } t \in [-1, 0) \text{ and } x(t) = x(T) \text{ for } t \in (T, T + 1]\}.
\]
Now, for any $x \in \tilde{D}_{[0,T]}^m$, $t \in [0,T]$, and $\delta > 0$ define, respectively, the oscillation function and the WM1-oscillation of $x$ around $t$ by

$$\bar{v}(x, t, \delta) := \sup_{-1 \leq t_1 \leq t_2 \leq (t+\delta) \wedge (T+1)} \left| x(t_1) - x(t_2) \right|,$$

$$w_w(x, t, \delta) := \sup_{-1 \leq t_1 \leq t_2 \leq (t+\delta) \wedge (T+1)} \left| x(t_2) - \left| x(t_1), x(t_3) \right| \right|,$$

where $|z - A|$ is the Euclidean distance between the point $z$ and a subset $A$ in $\mathbb{R}^m$. Also, set up

$$w_w(x, \delta) := \sup_{-1 \leq t \leq T+1} w_w(x, t, \delta).$$

Define also the set of discontinuities of $x$ by $\text{Disc}(x) := \{ t \in [0, T] : x(t) \neq x(t-) \}$. Finally, define the metric $d_p$ on $\tilde{D}_{[0,T]}^m$ by

$$d_p(x, y) = \max_{1 \leq i \leq m} d_w(x_i, y_i).$$

This is the metric that induces the product topology. From the second representation in (3.1) it follows that $d_p(x, y) \leq d_w(x, y)$. That is, the product topology is not stronger than the weak topology. The next theorem claims that the two topologies coincide. As a byproduct, we obtain that the weak topology is metrizable. The next proposition provides equivalent characterizations for the WM1 convergence.

**Proposition 3.1** (Theorem 12.5.2 in [30]) The following are equivalent characterizations of $x_n \to x$ as $n \to \infty$ in the WM1 topology of $\tilde{D}_{[0,T]}^m$.

(i) $d_w(x_n, x) \to 0$ as $n \to \infty$.

(ii) $d_p(x_n, x) \to 0$ as $n \to \infty$.

(iii) $x_n(t) \to x(t)$ as $n \to \infty$ for every $t \in [0, T] \setminus \text{Disc}(x)$ and

$$\lim_{\delta \to 0} \lim_{n \to \infty} \sup w_w(x_n, \delta) = 0.$$

(iv) $x_n(T) \to x(T)$ as $n \to \infty$; for every $t \in [0, T] \setminus \text{Disc}(x)$

$$\lim_{\delta \to 0} \lim_{n \to \infty} \sup_{-1 \leq s \leq (t+\delta) \wedge (T+1)} |x^n(s) - x(s)| = 0;$$

and for every $t \in \text{Disc}(x)$

$$\lim_{\delta \to 0} \lim_{n \to \infty} w_w(x_n, t, \delta) = 0.$$

**Corollary 3.1** Let $\{x^n\}_n \cup \{y^n\}_n \cup \{x\} \subset \tilde{D}_{[0,T]}^m$. If $d_w(x^n, x) \to 0$ and $y^n$ converges to $y \in C([0,T], \mathbb{R}^m)$ in the uniform norm, then $d_w(x^n + y^n, x + y) \to 0$. 

14
Proof. Part (iv) of the previous proposition holds true for \( \{x^n\}_n \) and \( x \). From the uniform convergence of \( y^n \) to \( y \) we clearly have \( y^n(T) \to y(T) \) and for every \( t \in [0, T] \),

\[
\lim_{\delta \to 0} \limsup_{n \to \infty} \sup_{-1 \leq s \leq (t+\delta) \wedge (T+1)} |y^n(s) - y(s)| = 0.
\]

Moreover, using that \( y \) is continuous, one gets that for every \( t \in \text{Disc}(x + y) = \text{Disc}(x) \),

\[
w_w(x_n + y_n, t, \delta) = w_w(x_n, t, \delta) + \varepsilon(n, \delta),
\]

where \( \lim_{\delta \to 0} \limsup_{n \to \infty} \varepsilon(n, \delta) = 0 \). Hence,

\[
\lim_{\delta \to 0} \limsup_{n \to \infty} w_w(x_n + y_n, t, \delta) = 0,
\]

and part (iv) of the previous proposition holds true for \( \{x^n + y^n\}_n \) and \( x + y \), hence, \( d_w(x^n + y^n, x + y) \to 0 \) as \( n \to \infty \).

Next, we provide a characterization of compactness.

**Proposition 3.2** (Theorem 12.12.2 in [36]) A subset \( A \) of \( \tilde{D}^m_{[0,T]} \) is relatively compact in the WM1 topology if

\[
\sup_{x \in A} \{|x|_T\} < \infty \tag{3.6}
\]

and

\[
\lim_{\delta \to 0} \sup_{x \in A} w_w(x, \delta) = 0.
\]

An important observation is that for every function \( x \in \tilde{D}^m_{[0,T]} \) with nondecreasing components in the sense that \( t \mapsto x_i(t) \) is nondecreasing for each \( 1 \leq i \leq m \), one has \( w_w(x, \delta) = 0 \). Hence for compactness, it is sufficient to verify only (3.6). This is summarized in the next corollary.

**Corollary 3.2** Fix \( K > 0 \). The set

\[
\left\{ x \in \tilde{D}^m_{[0,T]} : x_i \text{ is nondecreasing and } |x_i|_T \leq K, 1 \leq i \leq m \right\}
\]

is compact under the WM1 topology.

We end this section by establishing the connection to probability. For this, it is important to recall that the WM1 topology is metrizable by \( d_p \). Hence, Prohorov’s theorem applies and tightness is equivalent to relatively compactness, see [6, Theorems 5.1 and 5.2]

**Proposition 3.3** (Theorem 12.12.3 in [36]) A sequence \( \{\mathbb{P}^n\}_n \) of probability measures on \( \tilde{D}^m_{[0,T]} \) is tight under the WM1 topology if

\[
\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}^n \left( \left\{ x \in \tilde{D}^m_{[0,T]} : |x|_T > K \right\} \right) = 0 \tag{3.7}
\]

and for every \( \varepsilon > 0 \),

\[
\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}^n \left( \left\{ x \in \tilde{D}^m_{[0,T]} : w_w(x, \delta) \geq \varepsilon \right\} \right) = 0. \tag{3.8}
\]
4 Proof of Theorem 2.1 using the WM1 topology

From Corollary 3.2 it follows that the WM1 topology can handle quite easily nondecreasing processes (component-wise). Hence, in case that the processes $U$ and $\int_{[0,\cdot]} k(s)dU(s)$ are nondecreasing then (3.8) holds trivially and in order to establish compactness we only need to verify (3.7). These monotonocities follow if $U$ and $KU$ lie in the nonnegative orthant, where $K$ lies in the nonnegative orthant, where $KU$ lie in the nonnegative orthant, and $KU$ lie in the nonnegative orthant, where $K$ lies in the nonnegative orthant.

Similarly, there is a regular matrix $S$ so that by an invertible linear transformation, this cone can be mapped into the nonnegative orthant. From Corollary 3.2 it follows that the WM1 topology can handle quite easily nondecreasing processes (component-wise). Hence, in case that the processes $U$ and $\int_{[0,\cdot]} k(s)dU(s)$ are nondecreasing then (3.8) holds trivially and in order to establish compactness we only need to verify (3.7). These monotonocities follow if $U$ and $KU$ lie in the nonnegative orthant, where $K$ lies in the nonnegative orthant, where $KU$ lie in the nonnegative orthant, and $KU$ lie in the nonnegative orthant, where $K$ lies in the nonnegative orthant.

The observation is that, without loss of generality, we may assume the latter. Indeed, the second condition imposed in (2.1) implies that the closure of the convex cone $U$ lies in some open half space. By an invertible linear transformation, this cone can be mapped into the nonnegative orthant, see a two-dimensional example in Figure 4. Identify this linear transformation by an invertible matrix $S_1 \in \mathbb{R}^{d_1 \times d_1}$ and denote $U' := S_1U$. Then the process $U$ can be expressed by $U(t) = S_1^{-1}U'(t)$ for some process $U'$ with increments in the nonnegative orthant of $\mathbb{R}^{d_1}$.

Similarly, there is a regular matrix $S_2 \in \mathbb{R}^{d \times d}$, with finite norm $|S_2| := \sup_{\|x\| \in \mathbb{R}^d} |S_2x|/|x|$, such that $KU' = S_2KU$ lies in the nonnegative orthant, where $K := S_2K S_1^{-1}$.

Set, $X'(t) = S_2X(t)$, $b'(t, x) = S_2b(t, S_2^{-1}x)$, $\sigma'(t, x) = S_2\sigma(t, S_2^{-1}x)$, $k'(t) = S_2k(t)S_1^{-1}$.

All together, the new state process $X'$ satisfies

$$X'(t) = X'(0) + \int_0^t b'(s, X'(s))ds + \int_0^t \sigma'(s, X'(s))dW(s) + \int_{[0,t]} k'(s)dU'(s), \quad t \in [0,T].$$

Similarly, set up the new cost components

$$f'(t, x) = f(t, S_2^{-1}x), \quad g'(t, x) = g(t, S_2^{-1}x), \quad h'(t) = h(t)S_1^{-1},$$

and $X' := S_2X$. Assumption 2.1 clearly holds now with $u_1 = (1, \ldots, 1) \in \mathbb{R}^{d_1}$, $v_1 = (1, \ldots, 1) \in \mathbb{R}^d$, and $a_0 = 1$. Assumption 2.2 holds for the new components, where the Lipschitz continuity follows since $S_2$ has a finite norm. The invertibility of $S_1$ and $S_2$ enables to go back from the new problem to the original one. Also, notice that Assumption 2.2.4 is translated now to $h'(s) \geq 0$ component-wise, for every $s \in [0,T]$.

Another geometric way to look at it is that there are at most $d_1$ vectors in $\mathbb{R}^{d_1}$ such that any point in the cone $U$ is a linear combination of these vectors with nonnegative coefficients. These vectors point to ‘extreme’ directions (being at the boundary of the cone) of the jumps, and the coefficients, being nonnegative, mean that jumps always go in the same directions component-wise. For illustration, consider the cone $\{\alpha(-1, -1) + \beta(1/3, 1) : \alpha, \beta \geq 0\}$ from Figure 4. Each point in this cone is indeed a nonnegative combination of the (extreme) vectors $(1, -1)$ and $(1/3, 1)$.

Following the discussion above, in the rest of the proof we assume without loss of generality that

$$U \subseteq \mathbb{R}^{d_1}_+, \quad KU \subseteq \mathbb{R}^d_+, \quad h(s) \geq 0 \text{ for every } s \in [0,T]. \quad (4.1)$$

Hence,

$$t \mapsto \left(U(t), \int_{[0,t]} k(s)dU(s)\right) \text{ is non decreasing component-wise.}$$
In order to establish the existence of an optimal control, we start with a sequence of asymptotic optimal controls. Then we show that this sequence is relatively compact, hence has a converging subsequence and that any of its limit points is an optimal control (Proposition 4.1).

Specifically, we consider a sequence of admissible singular controls $\Xi^n = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,T]}, \mathbb{P}^n, X, W, U)$ such that

$$
\lim_{n \to \infty} J(\mathbb{P}^n) = V.
$$

Before arguing its tightness (under WM1), we prove that the control and the state processes have finite moments of orders $\bar{p}$ and $p$, respectively. These bounds serve us in the proofs of the next three Propositions. Let $\mathbb{E}^n = \mathbb{E}^{\mathbb{P}^n}$ be the expectation with respect to the measure $\mathbb{P}^n$.

**Lemma 4.1** The following two bounds hold:

$$
\sup_n \mathbb{E}^n[|U(T)|^{\bar{p}}] < \infty
$$

and

$$
\sup_n \sup_{t \in [0,T]} \mathbb{E}^n[|X(t)|^p] < \infty.
$$

**Proof.** We assume that part (a) of Assumption 2.2.5 holds. The proof in case that (b) holds is similar and therefore omitted. Throughout the proof $C$ refers to a positive constant, independent of $t$ and $n$, and which can change from one line to the next. Recall parts 2 and 3 in Assumption 2.2. Burkholder–Davis–Gundy (BDG) inequality implies that there exists a constant $C > 0$, such that for every $t \in [0,T]$ and $n \in \mathbb{N},$

$$
\mathbb{E}^n[|U(T)|^{\bar{p}}] \leq C(1 + \mathbb{E}^n[|X(T)|^p]), \quad \mathbb{E}^n[|X(t)|^p] \leq C\left(1 + \mathbb{E}^n[|U(t)|^p]\right).
$$

Figure 4: A linear transformation mapping the cone $\{\alpha(-1,-1) + \beta(1/3,1) : \alpha, \beta \geq 0\}$ to the nonnegative orthant.
Moreover, pay attention that the boundedness of \( h \) and the monotonicity of \( U \) imply that \( |\int_0^T h(s) dU(s)| \leq C|U(T)| \). Then, Assumption 2.2 and both parts of (4.5) imply the following two inequalities, respectively,

\[
J(\mathbb{P}^n) := \mathbb{E}^n \left[ \int_0^T f(t, X(t)) dt + g(X(T)) + \int_{[0,T]} h(t) dU(t) \right] \\
\geq -C \left( 1 + \int_0^T \mathbb{E}^n[|X(t)|^p] dt - \mathbb{E}^n[|X(T)|^p] + \mathbb{E}|U(T)| \right) \\
\geq -C \left( 1 + \int_0^T \mathbb{E}^n[|U(t)|^p] dt - \mathbb{E}^n[|U(T)|^p] + \mathbb{E}|U(T)| \right).
\]

Isolating \( \mathbb{E}^n[|U(T)|^\bar{p}] \) in the above, one gets that

\[
\mathbb{E}^n[|U(T)|^\bar{p}] \leq J(\mathbb{P}^n) + C \left( 1 + \mathbb{E}^n[|U(T)|^p] + \mathbb{E}^n[|U(T)|] \right) \\
\leq C \left( 1 + \mathbb{E}^n[|U(T)|^p] \right),
\]

where the first inequality follows since \( t \mapsto |U(t)| \) is nondecreasing, and the second inequality follows since by (4.2), \( \sup_n J(\mathbb{P}^n) < \infty \) and since \( p \geq 1 \). By an application of Young’s inequality and since \( \bar{p} > p \), it follows that (4.3) holds.

Keeping in mind that \( \bar{p} > p \), combining (4.3) with the second part of (4.5), we get that (4.4) holds as well.

\[ \square \]

Set the processes

\[
L(t) = \int_0^t b(s, X(s)) ds + \int_0^t \sigma(s, X(s)) dW(s), \\
Z(t) = \int_{[0,t]} G(s) dU(s).
\]

**Proposition 4.1** The sequence \( \mathbb{P}^n \circ (X, L, U, Z)^{-1} \) is relatively compact.

**Proof.** Recall that WM1 is metrizable (see Section 3.3). Hence Prohorov’s theorem holds and relatively compactness is equivalent to tightness. Pay attention that \( w_w((x, y), \delta) \leq w_w(x, \delta) + w_w(y, \delta) \). Hence, in order to establish the tightness of \( \mathbb{P}^n \circ (X, L, U, Z)^{-1} \) it is sufficient to show that each of the components in this sequence is tight. We start with the sequence of measures \( \{\mathbb{P}^n \circ (X)^{-1}\}_n \). To this end, it follows from Proposition 3.3 that it is sufficient to show that

\[
\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{P}^n(|X|_T > K) = 0,
\]

and that for every \( \varepsilon > 0 \),

\[
\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}^n(w_w(X, \delta) > \varepsilon) = 0. \tag{4.6}
\]

The first limit holds by Markov inequality and the following uniform bound,

\[
\sup_n \mathbb{E}^n[|X|^p_T] \leq C,
\]

18
which in turn follows by \([2.2]\), the boundedness of \(b\) and \(\sigma\), BDG inequality applied to \(\int_0^T \sigma \, dW(t)\), and \([4.3]\).

In order to establish \([4.6]\) pay attention first that the monotonicity of \(Z\) implies that for any \(t,t_1,t_2,t_3 \in [0,T]\), and \(\delta > 0\) satisfying \(0 < t < t_2 < (t + \delta) \land T\), one has

\[
|X(t_2) - [X(t_1),X(t_3)]| \leq \left| \int_{t_1}^{t_2} b(s,X(s))ds \right| + \left| \int_{t_1}^{t_2} \sigma(s,X(s))dW(s) \right| \\
+ \left| \int_{t_2}^{t_3} b(s,X(s))ds \right| + \left| \int_{t_2}^{t_3} \sigma(s,X(s))dW(s) \right|.
\]

Again, the boundedness of \(b\) and \(\sigma\) and BDG inequality imply that

\[
\mathbb{P}^n[(w_w(X,\delta))^2] \leq C(\delta + \delta^2),
\]

for some constant \(C > 0\), independent of \(n\) and \(\delta\). Markov inequality implies \([4.6]\). The tightness of \(\mathbb{P}^n \circ L^{-1}\) follows by the same arguments.

Recall Corollary \([3.2]\). The tightness of \(\mathbb{P}^n \circ (U,Z)^{-1}\) follows since \(k\) is bounded (see Assumption \([2.2]\)), \((U,Z)\) is nondecreasing component-wise, and from the bound \([4.3]\). Altogether, we obtain that \(\mathbb{P}^n \circ (X,L,U,Z)^{-1}\) is tight, and hence relatively compact.

\[\square\]

We now identify the limit points of \(\mathbb{P}^n \circ (X,L,U,Z)^{-1}\). Let \(\mathbb{P} \circ (X,L,U,Z)^{-1}\), be a limit point. By Skorokhod’s representation theorem and by reducing to a subsequence, which we relabel by \(\{n\}\), we may consider a probability space \((\Omega, \mathcal{G}, \mathbb{Q})\) that supports a sequence of processes \(\{(X^n,L^n,U^n,Z^n)\}_n\) and the processes \((\bar{X},\bar{L},\bar{U},\bar{Z})\), such that

\[
\mathbb{Q} \circ (X^n,Y^n,U^n,Z^n)^{-1} = \mathbb{P}^n \circ (X^n,Y^n,U^n,Z^n)^{-1},
\]

\[
\mathbb{Q} \circ (\bar{X},\bar{L},\bar{U},\bar{Z})^{-1} = \mathbb{P} \circ (X,L,U,Z)^{-1},
\]

and

\[
d_{\mathbb{P}}((X^n,L^n,U^n,Z^n), (\bar{X},\bar{Y},\bar{U},\bar{Z})) \to 0, \quad \mathbb{Q}\text{-a.s.}
\]

Specifically, \(\mathbb{Q}\)-almost surely (a.s.), for every \(t \in [0,T]\), \(X^n(t) = x + L^n(t) + Z^n(t)\), where

\[
L^n(t) = \int_0^t b(s,X^n(s))ds + \int_0^t \sigma(s,X^n(s))dW^n(s), \quad Z^n(t) = \int_{[0,t]} k(s)dU^n(s).
\]

Also, set the filtration

\[
\mathcal{G}_t := \sigma\{\bar{X}(s),\bar{L}(s),\bar{U}(s),\bar{Z}(s) : s \leq t\}.
\]

**Proposition 4.2** The processes \(\bar{X},\bar{L},\bar{U}\), and \(\bar{Z}\) satisfy \(\mathbb{Q}\)-a.s., for every \(t \in [0,T]\),

\[
\bar{Z}(t) = \int_{[0,t]} k(s)d\bar{U}(s),
\]

\[
\bar{L}(t) = \int_0^t b(s,\bar{X}(s))ds + \int_0^t \sigma(s,\bar{X}(s))d\bar{W}(s),
\]

\[
\bar{X}(t) = x + \bar{L}(t) + \bar{Z}(t),
\]

for some \(\mathcal{G}_t\)-Wiener process \(\bar{W}\). Furthermore, \(\mathbb{Q}\)-a.s., \(\bar{U}\) has increments in \(\mathcal{U}\), and \(X(t) \in \mathcal{X}\) for all \(t \in [0,T]\).
Proof. The convergence \[4.8\] implies the convergence of each of the components under the metric \(d_p\). The first observation is that by Proposition 3.1(iii) the convergence \(d_d(U^n, U) \to 0\) implies the convergence of \(U^n(t) \to \bar{U}(t)\) for any continuity point of \(U^n\) and \(U^n(T) \to \bar{U}(T)\). Therefore, the portmanteau theorem [6, Theorem 2.1] implies that \[4.10\] holds.

The proof of \[4.11\] uses standard martingale arguments. However, for completeness of the presentation we provide the details. For every infinity differentiable function \(\phi : \mathbb{R}^d \to \mathbb{R}\), define the generator

\[
G\phi(t, x) := b(t, x) \cdot D\phi(x) + \frac{1}{2} \text{Tr} \left[ \sigma \sigma^T(t, x) D^2 \phi(x) \right],
\]

where \(D\phi\) and \(D^2\phi\) are, respectively, the gradient and the Hessian of \(\phi\). Also, set the processes

\[
M^n_\phi(t) := \phi(L^n(t)) - \int_0^t G\phi(s, X^n(s))ds, \quad t \in [0, T],
\]

\[
M_\phi(t) := \phi(\bar{L}(t)) - \int_0^t G\phi(s, \bar{X}(s))ds, \quad t \in [0, T].
\]

Fix an arbitrary infinity differentiable function \(\phi\) with compact support. Once we show that \(M_\phi\) is a \(G_t\)-martingale, it follows that \(\bar{L}(t) = \int_0^t b(s, \bar{X}(s))ds + \int_0^t \sigma(s, \bar{X}(s))d\bar{W}(s)\), for some \(G_t\)-Wiener process \(\bar{W}\). Fix arbitrary \(s, t > 0\) and a continuous function \(F^s\) on its domain with respect to the WM1 topology. Then,

\[
0 = \mathbb{E}^Q \left[ F^s(X^n(u), L^n(u), U^n(u), Z^n(u); u \leq s)(M^n_\phi(t) - M^n_\phi(s)) \right]
\]

\[
\to \mathbb{E}^Q \left[ F^s(\bar{X}(u), \bar{L}(u), \bar{U}(u), \bar{Z}(u); u \leq s)(M_\phi(t) - M_\phi(s)) \right],
\]

as \(n \to \infty\). By the definition of \(G\) it follows that \(M_\phi\) is a \(G_t\)-martingale.

Notice that \(\bar{L}\) is continuous, hence Corollary 3.1 together with \[4.8\] and \[4.9\] imply that \(\bar{X}(t) = x + \bar{L}(t) + \bar{Z}(t)\).

Finally, the control and state constraints \(\bar{X}(t) \in \mathcal{X}\) and \(U(t) - U(s) \in \mathcal{U}, 0 \leq s \leq t \leq T\), hold \(\mathbb{Q}\)-a.s. since they hold in the prelimit and the sets \(\mathcal{X}\) and \(\mathcal{U}\) are closed.

The next proposition points out an optimal control, hence establishing Theorem 2.1.

**Proposition 4.3** The control \(\Xi := (\Omega, \bar{G}, (\bar{G}_t)_{t \in [0, T]}, \mathbb{Q}, \bar{X}, \bar{W}, \bar{U})\) is optimal.

**Proof.** From Proposition 4.2 it follows that \(\Xi\) is admissible. Hence, we only need to show that its associated cost equals \(\bar{V}\). By \[4.2\] and \[4.7\] it is sufficient to show that

\[
\lim_{n \to \infty} \mathbb{E}^Q \left[ \int_0^T f(t, X^n(t))dt + g(X^n(T)) + \int_{[0, T]} h(s)dU^n(s) \right] = \mathbb{E}^Q \left[ \int_0^T f(t, \bar{X}(t))dt + g(\bar{X}(T)) + \int_{[0, T]} h(s)d\bar{U}(s) \right].
\]

The same arguments leading to \[4.10\] imply that

\[
\lim_{n \to \infty} \mathbb{1}_{\{|U^n(T)| \leq M\}} \int_{[0, T]} h(s)dU^n(s) = \mathbb{1}_{\{|\bar{U}(T)| \leq M\}} \int_{[0, T]} h(s)d\bar{U}(s), \quad \mathbb{Q}\text{-a.s., under } d_p,
\]
Recall (4.1), that is, $U \subseteq \mathbb{R}^d_+$ and $h \geq 0$. Moreover, by Assumption (2.2), $h$ is bounded. Hence, the bounded convergence theorem implies that
\[
\lim_{n \to \infty} \mathbb{E}^Q \left[ \mathbb{1}_{\{|U^n(T)| \leq M\}} \int_{[0,T]} h(s) dU^n(s) \right] = \mathbb{E}^Q \left[ \mathbb{1}_{\{|U(T)| \leq M\}} \int_{[0,T]} h(s) dU(s) \right].
\]

Moreover, the monotone convergence theorem implies that
\[
\lim_{M \to \infty} \mathbb{E}^Q \left[ \mathbb{1}_{\{|U(T)| \leq M\}} \int_{[0,T]} h(s) d\bar{U}(s) \right] = \mathbb{E}^Q \left[ \int_{[0,T]} h(s) d\bar{U}(s) \right].
\]

Pay attention that
\[
\mathbb{E}^Q \left[ \mathbb{1}_{\{|U^n(T)| > M\}} \int_{[0,T]} h(s) dU^n(s) \right] = \mathbb{E}^Q \left[ \int_{[0,T]} h(s) dU^n(s) \right] - \mathbb{E}^Q \left[ \mathbb{1}_{\{|U^n(T)| \leq M\}} \int_{[0,T]} h(s) dU^n(s) \right].
\]

Since $U$ is nondecreasing and $h$ is bounded (see Assumption 2.2.1) it follows from Lemma 4.1 that
\[
\lim_{M \to \infty} \sup_n \mathbb{E}^Q \left[ \int_{[0,T]} h(s) dU^n(s) \right] = \mathbb{E}^Q \left[ \int_{[0,T]} h(s) d\bar{U}(s) \right] = 0.
\]

Combining (4.13), (4.14), and (4.15), we obtain the convergence,
\[
\lim_{n \to \infty} \mathbb{E}^Q \left[ \int_{[0,T]} h(s) dU^n(s) \right] = \mathbb{E}^Q \left[ \int_{[0,T]} h(s) d\bar{U}(s) \right].
\]

Recall the growth assumptions on $f$ and $g$ and their continuity. Repeating the same arguments leading to the last limit, where now using the $\mathbb{Q}$-a.s. convergence $d_p(X^n, X) \to 0$, (4.4), and truncating $f$ and $g$ by $M > 0$, we get that
\[
\lim_{n \to \infty} \mathbb{E}^Q \left[ \int_0^T f(t, X^n(t)) dt + g(X^n(T)) \right] = \mathbb{E}^Q \left[ \int_0^T f(t, X(t)) dt + g(X(T)) \right].
\]

**References**

[1] S. Asmussen and M. Taksar. Controlled diffusion models for optimal dividend pay-out. *Insurance Math. Econom.*, 20(1):1–15, 1997.

[2] R. Atar and A. Budhiraja. Singular control with state constraints on unbounded domain. *Ann. Probab.*, 34(5):1864–1909, 2006.

[3] R. Atar and M. Shifrin. An asymptotic optimality result for the multiclass queue with finite buffers in heavy traffic. *Stoch. Syst.*, 4(2):556–603, 2014.

[4] B. Basrak and D. Krizmanić. A multivariate functional limit theorem in weak m1 topology. *J. Theoret. Probab.*, 28(1):119–136, Mar 2015.

[5] J. Bather and H. Chernoff. Sequential decisions in the control of a spaceship. In *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 3, pages 181–207, 1967.
[6] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[7] A. Budhiraja and A. P. Ghosh. Diffusion approximations for controlled stochastic networks: an asymptotic bound for the value function. *Ann. Appl. Probab.*, 16(4):1962–2006, 2006.

[8] A. Budhiraja and K. Ross. Existence of optimal controls for singular control problems with state constraints. *Ann. Appl. Probab.*, 16(4):2235–2255, 2006.

[9] A. Budhiraja and K. Ross. Convergent numerical scheme for singular stochastic control with state constraints in a portfolio selection problem. *SIAM J. Control Optim.*, 45(6):2169–2206, Jan. 2007.

[10] A. Cohen. Asymptotic analysis of a multiclass queueing control problem under heavy-traffic with model uncertainty. *Stoch. Syst.*, to appear, 2019.

[11] M. H. Davis and A. R. Norman. Portfolio selection with transaction costs. *Math. Oper. Res.*, 15(4):676–713, 1990.

[12] F. Delarue, J. Inglis, S. Rubenthaler, and E. Tanr. Particle systems with a singular mean-field self-excitation. Application to neuronal networks. *Stochastic Process. Appl.*, 125(6):2451–2492, 2015.

[13] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.

[14] G. Fu and U. Horst. Mean field games with singular controls. *SIAM J. Control Optim.*, 55(6):3833–3868, 2017.

[15] J. M. Harrison. Brownian models of queueing networks with heterogeneous customer populations. In *Stochastic differential systems, stochastic control theory and applications (Minneapolis, Minn., 1986)*, volume 10 of *IMA Vol. Math. Appl.*, pages 147–186. Springer, New York, 1988.

[16] J. M. Harrison and M. I. Taksar. Instantaneous control of Brownian motion. *Math. Oper. Res.*, 8(3):439–453, 1983.

[17] J. M. Harrison and J. A. Van Mieghem. Dynamic control of Brownian networks: state space collapse and equivalent workload formulations. *Ann. Appl. Probab.*, 7(3):747–771, 1997.

[18] J. M. Harrison and R. J. Williams. A multiclass closed queueing network with unconventional heavy traffic behavior. *Ann. Appl. Probab.*, 15(4):2255–2295, 2005.

[19] J. M. Harrison and R. J. Williams. Workload reduction of a generalized Brownian network. *Ann. Appl. Probab.*, 6(1):1–47, 1996.

[20] O. Kella and W. Whitt. Diffusion approximations for queues with server vacations. *Adv. in Appl. Probab.*, 22(3):706–729, 1990.

[21] T. G. Kurtz. Random time changes and convergence in distribution under the Meyer-Zheng conditions. *Ann. Probab.*, 19(3):1010–1034, 1991.

[22] H. J. Kushner and L. F. Martins. Numerical methods for stochastic singular control problems. *SIAM J. Control Optim.*, 29(6):1443–1475, 1991.

[23] A. Mandelbaum and W. A. Massey. Strong approximations for time-dependent queues. *Math. Oper. Res.*, 20(1):33–64, 1995.

[24] L. F. Martins and H. J. Kushner. Routing and singular control for queueing networks in heavy traffic. *SIAM J. Control Optim.*, 28(5):1209–1233, 1990.

[25] P.-A. Meyer and W. A. Zheng. Tightness criteria for laws of semimartingales. *Ann. Inst. H. Poincaré Probab. Statist.*, 20(4):353–372, 1984.

[26] S. Nadtochiy and M. Shkolnikov. Particle systems with singular interaction through hitting times: Application in systemic risk modeling. *Ann. Appl. Probab.*, 29(1):89–129, 2019.

[27] G. Pang, W. Whitt, et al. Continuity of a queueing integral representation in the m1 topology. *Ann. Appl. Probab.*, 20(1):214–237, 2010.

[28] A. A. Puhalskii and W. Whitt. Functional large deviation principles for first-passage-time processes. *Ann. Appl. Probab.*, 7(2):362–381, 1997.
[29] A. A. Puhalskii and W. Whitt. Functional large deviation principles for waiting and departure processes. *Probab. Engrg. Inform. Sci.*, 12(4):479–507, 1998.

[30] S. Resnick and E. van den Berg. Weak convergence of high-speed network traffic models. *J. Appl. Probability*, 37(2):575597, 2000.

[31] S. E. Shreve and H. M. Soner. A free boundary problem related to singular stochastic control. In *Proceedings of Imperial College Workshop on Applied Stochastic Analysis*. Citeseer, 1990.

[32] A. V. Skorokhod. Limit theorems for stochastic processes. *Theory Probab. Appl.*, 1(3):261–290, 1956.

[33] H. M. Soner and S. E. Shreve. Regularity of the value function for a two-dimensional singular stochastic control problem. *SIAM J. Control Optim.*, 27(4):876–907, 1989.

[34] W. Whitt. Weak convergence of first passage time processes. *J. Appl. Probability*, 8(2):417–422, 1971.

[35] W. Whitt. Limits for cumulative input processes to queues. *Probab. Engrg. Inform. Sci.*, 14(2):123–150, 2000.

[36] W. Whitt. *Stochastic-process limits*. Springer Series in Operations Research. Springer-Verlag, New York, 2002. An introduction to stochastic-process limits and their application to queues.

[37] M. J. Wichura et al. Functional laws of the iterated logarithm for the partial sums of iid random variables in the domain of attraction of a completely asymmetric stable law. *Ann. Probab.*, 2(6):1108–1138, 1974.