
In mechanics, nonholonomic constraints are constraints depending on velocities that cannot be reduced by integration to a set of constraints depending on positions only. A typical example of nonholonomic constraints is given by rolling without slipping. Nonholonomic dynamics is devoted to the study of motion of such (uncontrolled) mechanical systems [12]. However, in many applications, especially in robotics (e.g., robotic cars with trailers), one has the possibility of controlling the motion of such systems. This leads to nonholonomic control systems. The main objective of the book under review is to introduce the readers to nonholonomic systems from the point of view of control theory.

As a rule, nonholonomic constraints depend linearly in velocities, i.e., are described by systems of Pfaffian equations. The corresponding control systems depend linearly on control, i.e., they have the following form:

\[ \dot{q} = \sum_{i=1}^{m} u_i X_i(q), \quad q \in M, \ u = (u_1, \ldots, u_m) \in \mathbb{R}^m, \]

where \( M \) is a connected \( n \)-dimensional manifold, \( X_1, \ldots, X_m \) are smooth vector fields on \( M \), and \( m < n \).

The system (1) (with a fixed initial condition \( q(0) = q_0 \)) is controlled by choosing values \( u(t) \in \mathbb{R}^m \) of the control parameter at (almost) every time instant \( t \), where the function \( u(\cdot) \) belongs to an appropriate functional space that ensures the existence and uniqueness of the solution of the corresponding initial value problem.

The resulting trajectory is called an admissible trajectory of the control system (1) corresponding to the control function \( u(t) \) and the initial point \( q_0 \). The linear span \( \Delta(q) \) of the vectors \( X_1(q), \ldots, X_m(q) \) is called the set of admissible velocities of (1) at the point \( q \).

Admissible trajectories of system (1) are quite special: for almost every \( t \) the velocity at \( t \) of an admissible trajectory \( q(t) \) must belong to the set of admissible velocities \( \Delta(q(t)) \), which are proper subspaces of the tangent space \( T_{q(t)} M \). The first natural question about system (1) is: What is the reachable set from a given point \( q_0 \), i.e., the set \( R_{q_0} \) of points that can be reached from the point \( q_0 \) by moving along all admissible trajectories?

By the Nagano–Sussmann–Stefan orbit theorem [14][15] the reachable set \( R_{q_0} \) (with piecewise constant control functions) is an immersed submanifold of \( M \). The dimension of the reachable set can be (and generically) is bigger than the dimension of the set of admissible velocities (at a generic point), and the reason for this is that the flows generated by two vector fields usually do not commute. This noncommutativity is very well demonstrated by the parallel parking of a car, when a driver uses commutators of the rotation of the steering wheel and a linear motions of the car.

2010 Mathematics Subject Classification. Primary 70E60, 70Q05, 53C17, 57R27, 93C10.
Applying the commutators of the flows generated by two vector fields $X$ and $Y$ tangent to $\Delta$ at a point $q_0$, one moves infinitesimally in the direction of their Lie bracket $[X,Y](q_0)$. Therefore, if $\text{Lie}(X_1,\ldots,X_m)$ is the Lie algebra of vector fields generated by the vector fields $X_1,\ldots,X_m$ and $\text{Lie}(X_1,\ldots,X_m)(q_0)$ is the space of all vectors obtained by the evaluation of the elements of $\text{Lie}(X_1,\ldots,X_m)$ at $q_0$, then the dimension of the reachable set $\mathcal{R}_{q_0}$ is greater than or equal to the dimension of the vector space $\text{Lie}(X_1,\ldots,X_m)(q_0)$,

$$\dim \mathcal{R}_{q_0} \geq \dim \text{Lie}(X_1,\ldots,X_m)(q_0),$$

and the equality in (2) occurs in the real analytic category (i.e., when the ambient manifold $M$ and the vector fields $X_1,\ldots,X_m$ are real analytic) and if $\text{Lie}(X_1,\ldots,X_m)$ is a (locally) finitely generated module over $C^\infty(M)$ \cite{2,10}. The latter happens, for example, if $\dim \text{Lie}(X_1,\ldots,X_m)(q)$ is independent of $q$, i.e., if the family of subspaces $\{\text{Lie}(X_1,\ldots,X_m)(q)\}_{q \in M}$ defines a distribution on $M$ or, in other words, a subbundle of the tangent bundle $TM$.

There are the following two extreme cases:

1. **Holonomic (integrable, involutive, Frobenius) case.** Assume that $\Delta$ is a distribution of rank $m$ and $\text{Lie}(X_1,\ldots,X_m)(q) = \Delta(q)$ for all $q \in M$. The latter condition means that the Lie brackets of any two vector fields tangent to the distribution $\Delta$ are also tangent to the distribution $\Delta$, or, symbolically, if $\Delta^2(q) := \Delta(q) + [\Delta, \Delta](q)$, then $\Delta^2(q) = \Delta(q)$. In this case the distribution $\Delta$ is called involutive. As a consequence of the results stated in the previous paragraph and applied in this situation, one gets the classical Frobenius theorem: if $\Delta$ is an involutive distribution of rank $m$, then it admits a foliation of $m$-dimensional integral submanifolds of $\Delta$, i.e., submanifolds with tangent space at any point equal to the fiber of $\Delta$ at the same point. In this case the constraints given by the system (1) can be replaced by constraints depending on positions only, i.e., they are holonomic in the sense of the first sentence of this review.

2. **Completely nonholonomic (bracket generating, Rashevsky–Chow, Hörmander) case.** Assume that

$$\text{Lie}(X_1,\ldots,X_m)(q) = T_q M, \quad \forall q \in M.$$  

Then from (2) and connectivity of $M$ it follows that $\mathcal{R}_q = M$, i.e., one gets the classical Rashevsky–Chow theorem \cite{5,13}: under condition (3) (also called the Rashevsky–Chow, bracket-generating, or Lie algebra rank condition) any two points of $M$ can be connected by a piecewise smooth admissible trajectory of (1) or, equivalently, the system (1) is controllable by piecewise smooth trajectories. In this case no constraints depending on position only can be obtained from the constraints given by the system (1), which justifies the name “completely nonholonomic”. Note that in the theory of PDEs the same condition (3) is called the Hörmander condition, and it guarantees the hypoellipticity of second-order differential operators of type $\sum_{i=1}^m X_i^2$ or $\sum_{i=2}^m X_i^2 + X_1 + \cdots$, where “…” denotes terms of order smaller or equal than 1 \cite{8}.

Note that the intermediate cases between holonomic one and completely nonholonomic one can be reduced to the completely nonholonomic case under assumption that $\dim \text{Lie}(X_1,\ldots,X_m)(q)$ is independent of $q$ (and greater than $m$) if one replaces $M$ by a connected component of the integral submanifold of the (involutive) distribution $\{\text{Lie}(X_1,\ldots,X_m)(q)\}_{q \in M}$. In this case some (but not all) constraints given by the system (1) can be replaced by constraints depending on position only.
So, it is not so restrictive to assume from the beginning that system (1) is completely nonholonomic.

The Rashevsky–Chow condition ensures the existence of an admissible trajectory connecting two given points of $M$. The next natural task, obviously important for applications, is to solve the **motion planning problem**: namely, given two points $q_0$ and $q_1$ in $M$, find a control function which steers a control system (1) from $q_0$ to $q_1$ or, in other words, such that the corresponding trajectory of (1) starting at $q_0$ will terminate at $q_1$. Some proofs of the Rashevsky–Chow theorem (including the proof given in the book under review) are almost constructive in a sense that they provide an explicit class of control functions such that if one uses them with a given initial point $q_0$, the set of all terminal points of the corresponding admissible trajectories will sweep out a neighborhood of $q_0$. However, since all proofs rely on the implicit function theorem, they are still not satisfactory for solving the motion planning problem.

The solution of the nonholonomic motion planning problem consists mainly of the following three natural steps:

**Step 1.** Define properly the “first approximation” of system (1) which preserves the main anisotropic properties of this system and, in particular, the controllability property. The naive “linearization” of system (1) by “freezing” the vector fields $X_1, X_2, \ldots, X_m$ at the point $q_0$ and considering the corresponding system on $T_{q_0}M$ is not a good choice for the first approximation because the resulting system is involutive.

The true first approximation of system (1) at $q_0$, under certain regularity assumption that will be described later, is a special system of type

$$
\dot{q} = \sum_{i=1}^{m} u_i \tilde{X}_i(q), \quad q \in G, u = (u_1, \ldots, u_m) \in \mathbb{R}^m,
$$

where $G$ is a certain nilpotent Lie group such that the associated Lie algebra $\mathfrak{g}$ is graded, $\mathfrak{g} = \bigoplus_{j=-\mu}^{-1} \mathfrak{g}^j$, and it is generated by its component $\mathfrak{g}^{-1}$ (such a group $G$ is also called a Carnot group), while the vector fields $\tilde{X}_1, \ldots, \tilde{X}_m$ are left invariant on $G$ and they span $\mathfrak{g}^{-1}$. System (4) is called the nilpotent approximation at $q_0$ in control theory and (Tanaka) flat homogeneous structure (model) in nilpotent differential geometry developed by N. Tanaka and his school [16]. It plays a crucial role not only in the motion planning problem and other aspects of control theory, but also in PDEs (subelliptic equations and nonholonomic diffusion equations) and in the equivalence problem for filtered structures in differential geometry (which is the same as the problem of state-feedback equivalence of the corresponding control systems). An example of a Carnot group is the Heisenberg group, and it appears as an ambient manifold of systems (1) with the underlying distribution $\Delta$ being contact.

In general, without the regularity assumption, the nilpotent approximation is a special control system on a homogeneous space of a Carnot group, and it will be described in some more detail later.

**Step 2.** Try to solve explicitly the motion planning problem for the systems which appear as nilpotent approximations for system (1). The crucial point here, which, in my opinion, is the main contribution of the book (based on the recent paper [4]) is that every such nilpotent approximation can be lifted by a procedure, called
desingularization}, to the corresponding left-invariant system on the Carnot group with the Lie algebra being the free graded Lie algebra with \( m \) generators truncated from a certain weight, and for such systems the motion planning problem can be solved explicitly via step-by-step use of sinusoidal controls with sufficiently big integer frequencies, generalizing the well-known algorithm of Murray and Sastry for chain form systems \([11]\).

Step 3. Try to exploit the first two steps to obtain an approximated solution for the original motion planning problem. One possibility, discussed in the book in detail, is to use a Newton-like iterative procedure to find an approximated solution. More precisely, restricting ourselves to a coordinate neighborhood, one can replace the ambient manifold \( M \) by \( \mathbb{R}^n \). At each iterative step, we replace our system by its nilpotent approximation at the terminal point of the previous step (that now is considered as the initial one) with coordinates compatible in a natural way to the standard coordinates in \( \mathbb{R}^n \). Further, find the control function that steers the obtained nilpotent approximation from the terminal point of the previous iterative step to the terminal point of the original motion planning problem by the method of Step 2. Applying the same control function to steer the original control system from the terminal point of the previous iterative step, we arrive to a new point. Then repeat the same procedure to this new point until we obtain a point which is sufficiently close to the terminal point of the original motion planning problem.

The rest of this review is mainly devoted to the discussion of the construction of the nilpotent approximation and the desingularization procedure. There are several equivalent ways to construct the nilpotent approximation; most of them are discussed in some form in the book. The primary approach of the book is via the construction of special local coordinates, the privileged coordinates, and the exposition follows closely the influential paper of A. Bellaïche \([3]\). My preferred way is to use pure Lie algebraic description (following N. Tanaka \([16]\) in the regular case and based on the ideas of S. Ignatovich \([9]\) in the general case). This way is short enough (even in order to include it in this review in a self-contained manner), a priori intrinsic, and does not use any auxiliary tool such as privileged coordinates and/or the (generalized) sub-Riemannian metric associated with system \([1]\) (the metric tangent space approach in the Gromov–Hausdorff sense \([6]\)). Besides, the desingularization procedure of Step 2 is already built into this Lie algebraic approach, as will be seen later. Note that an alternative coordinate-free approach to nilpotent approximation can be found in \([1]\).

To describe the Lie algebraic approach under regularity assumptions, we first construct the corresponding graded Lie algebra \( \mathfrak{g}(q_0) \) by passing from the natural filtration generated by the distribution \( \Delta \) on the tangent space \( T_{q_0}M \) to the corresponding graded space, and then from the Lie algebra \( \mathfrak{g} \) to the corresponding connected and simply connected Lie group \( G \).

More precisely, first by taking iterative Lie brackets of vector fields tangent to a distribution \( \Delta \), one can define a filtration

\[
\Delta^1 \subset \Delta^2 \subset \cdots \subset T_q M
\]

of the tangent bundle \( TM \), called a weak derived flag or a small flag (of \( \Delta \)). For this set \( \Delta = \Delta^1 \), and let \( \Delta^j(q) \) be the linear span of all iterative brackets of vector fields \( X_1, X_2, \ldots, X_m \) up to the length \( j \) evaluated at the point \( q \). The Rashevsky–Chow condition is equivalent to the fact that for every point \( q \in M \) there exists a positive
integer $\mu(q)$, called the degree of nonholonomy of the distribution $\Delta$ at $q$, such that $\Delta^{\mu(q)}(q) = T_qM$. The point $q_0$ is a regular point of the distribution $\Delta$ if for every positive $j$ the dimension of subspaces $\Delta^j(q_0)$ is constant in a neighborhood of $q_0$, and it is called singular otherwise. Let $g^{-1}(q) \overset{def}{=} \Delta(q)$ and $g^j(q) \overset{def}{=} \Delta^{-j}(q)/\Delta^{-j-1}(q)$ for $j < -1$. It turns out that for a regular point $q_0$ the degree of nonholonomy $\mu$, the graded space

$$g(q_0) = \bigoplus_{j=-\mu}^{-1} g^j(q_0),$$

corresponding to the filtration $[3]$ is endowed naturally with the structure of a graded nilpotent Lie algebra, generated by $g^{-1}(q_0)$. Informally speaking, this Lie algebra contains the information about the principal components of all iterative Lie brackets at $q_0$ of vector fields tangent to $\Delta$. Further, let $G$ be the connected, simply connected group with Lie algebra $g$. Note that under the identification of $g(q_0)$ with the tangent space $T_eG$ to $G$ at the identity $e$ and the fact that $\Delta(q_0) = g^{-1}(q_0) \subset g(q_0) \overset{def}{=} T_eG$, all vectors $X_i(q_0)$, $1 \leq i \leq m$, belong to $T_eG$. Finally, let $\hat{X}_i$ be the left-invariant vector field on $G$ such that $\hat{X}_i(e) = X_i(q_0)$. Then system $[4]$ with the constructed vector fields $\hat{X}_i$ is called the nilpotent approximation of the system $[1]$ at the regular point $q_0$.

If $q_0$ is not regular, consider the free Lie algebra $f_m$ with $m$ generators. The algebra $f_m$ has natural grading $f_m = \bigoplus_{i \in \mathbb{Z}_{-}} f^i_m$, where $\mathbb{Z}_{-}$ denotes the set of negative integers, $f^{-1}_m$ is the linear span of a (minimal) set of generators of $f_m$ and more generally, $f^i_m$ is the linear span of all brackets of length $i$ of a set of generators. Fix a set of generators $\{\ell_1, \ldots, \ell_m\}$ of $f_m$. There exists the unique Lie algebra isomorphism $\Phi : f_m \to \text{Lie}(X_1, \ldots, X_m)$ sending $\ell_i$ to $X_i$ for any $i = 1, \ldots, m$. Now fix a point $q_0 \in M$, and let $\Psi : f_m \to T_{q_0}M$ be the following map: $\Psi(\ell) = \Phi(\ell)(q_0)$ for any $\ell \in f_m$. In other words, $\Psi(\ell)$ is the evaluation of the vector field $\Phi(\ell)$ at the point $q_0$. For any $i \in \mathbb{Z}_{-}$, let $h^i$ be the following subspace of $f^i_m$:

$$h^i = \{ \ell \in f^i_m : \Psi(\ell) \in \Delta^{-i-1}(q_0) \}.$$

Let $h = \bigoplus_{i \in \mathbb{Z}_{-}} h^i$. Although the maps $\Phi$ and $\Psi$ depend on the choice of the generators $\{\ell_1, \ldots, \ell_m\}$ of $f_m$, the subspace $h$ is independent of this choice. Moreover, it is a graded subalgebra of $f_m$. Following [2], we call $h$ the core algebra of the generalized distribution $D$ at the point $q_0$.

How can one construct the nilpotent approximation of system $[1]$ from its core algebra $h$? Since the system has a finite degree of nonholonomy $\mu = \mu(q_0)$ at $q_0$, we can replace $f_m$ and $h$ by their truncated finite-dimensional parts $f_{m,\mu}$ and $h_{\mu}$ by removing all components of weight greater than $\mu(q_0)$. Let $F_{m,\mu}$ and $H$ be the connected and simply connected Lie groups with the Lie algebras $f_{m,\mu}$ and $h_{\mu}$, respectively. Further, let $L_i$ be the left-invariant vector field on the Lie group $F_{m,\mu}$ equal to $\ell_i$ at the identity of $F_{m,\mu}$. Let $F_{m,\mu}/H = \{ Hg : g \in F_{m,\mu} \}$ be the set of all right cosets of the Lie group $H$. If we denote by $\Pi : F_{m,\mu} \to F_{m,\mu}/H$ the canonical projection, set $\hat{X}_i = \Pi\ell_i$. Then system $[1]$ with the constructed vector fields $\hat{X}_i$ is called the nilpotent approximation of system $[1]$ at the regular point $q_0$.

Why, in the case of regular point $q_0$, does this construction give the same as in the previous one? The reason is that in the regular case the core algebra $h$ is an ideal of $f_m$, and so $G = F_{m,\mu}/H$ is a Lie group.
The desingularization procedure is nothing but the lifting of system (4) to the following left-invariant system on the group $F_{m,\mu}$:

\[
\dot{Q} = \sum_{i=1}^{m} u_i L_i(Q), \quad Q \in F_{m,\mu}, u = (u_1, \ldots, u_m) \in \mathbb{R}^m.
\]

In this way, if $q_0$ is a singular point of the degree of nonholonomy $\mu$ for the original system, then system (7), corresponding to the lift of its nilpotent approximation at $q_0$, consists of regular points of the degree of nonholonomy $\mu$ only. Note that this desingularization procedure is not relevant in the equivalence problem, because nonequivalent distributions of the same rank have the same desingularization. However, this procedure is quite useful for motion planning, because, informally speaking, one can project the trajectory which solves a steering problem for the lifted system to the original one. Besides, it is much simpler to deal with the set of lifted systems (7) than with the set of all possible nilpotent approximations (4) because the former set is discrete, and each system (7) can be written in a convenient way in appropriate coordinates (Hall–Grayson–Grossmann normal form [7]), while the set of all nilpotent approximations (4) cannot be explicitly classified and depends on continuous parameters.

The method of privileged coordinates, taken as the primary one in the book, is the most popular method for construction of a nilpotent approximation in control theory literature, maybe due to its elementary nature. It also can be seen in essence as the coordinate realization of the Lie algebraic approach: one can define a nonholonomic order of a smooth function $f$ at a point $q_0$ as the biggest integer $k$ such that for any $l < k$

\[
X_1 \circ X_2 \circ \cdots \circ X_l(f)(q_0) = 0
\]

for any $l$ vector fields $X_1, \ldots, X_l$ tangent to $\Delta$ (here the vector fields are considered as derivation of the algebra of smooth functions). The nonholonomic order on the algebra of functions induces a nonholonomic order of vector fields at a point in a natural way. A system of local coordinates $(x_1, x_2, \ldots, x_n)$ is called privileged if the first $m$ of the $x_i$’s have nonholonomic order at $q_0$ equal to 1, the next $\dim \mathfrak{g}^{-2}(q_0)$ $x_i$’s have the nonholonomic order at $q_0$ equal to 2, the next $\dim \mathfrak{g}^{-3}(q_0)$ $x_i$’s have the nonholonomic order at $q_0$ equal to 3, etc., where spaces $\mathfrak{g}^i(q_0)$ are as in (6). There are several standard ways to construct privileged coordinates: e.g., coordinates of first and second kind with respect to the specially adapted frame and more effective “algebraic” constructions, which do not require integrating flows of vector fields. Fixing a system of privileged coordinates, one gets a natural splitting of the space of vector fields into homogeneous components (with respect to the nonholonomic order at $q_0$). In particular, replacing the vector fields $X_i$ in system (1) by their homogeneous component of the minimal possible degree $-1$, one again obtains the nilpotent approximation of (1).

Finally, another way to understand the nilpotent approximation of system (1) is by introducing the (generalized) sub-Riemannian structure: if the vector fields $X_1, \ldots, X_m$ from (1) are linearly independent at every point, then there exists a unique Euclidean structure on each fiber of the corresponding distribution $\Delta$ for which the vector fields constitute an orthonormal frame. The distribution $\Delta$ endowed with Euclidean structure on each fiber is called the sub-Riemannian structure on $M$ with the underlying distribution $D$. Once a sub-Riemannian structure is given, one can define the length of any curve tangent to the distribution $\Delta$. If the
distribution is completely nonholonomic, the sub-Riemannian distance between any two points is the infimum of the lengths of all admissible curves connecting these two points, which in turn defines the sub-Riemannian metric space. In general, by topological reasons, the vector fields might be dependent at some points. In this case one can generalize the sub-Riemannian distance by assuming that we consider only admissible trajectories of system (1) with the control functions $u(\cdot)$ taking values on the unit sphere of $\mathbb{R}^m$. The generalized sub-Riemannian distance $d_{SR}(q_0, q_1)$ between two points $q_0$ and $q_1$ is the infimum of time required to steer from $q_0$ to $q_1$ moving only along the admissible trajectories satisfying this additional property.

With this metric point of view, one can further clarify many previous constructions starting from the notion of the nonholonomic order of a function $f$ at a point $q_0$, which turns to be exactly the integer $k$ such that $f(p) = O(d_{SR}(p, q_0)^k)$, and end up with the purely metric construction of the nilpotent approximation itself: the generalized sub-Riemannian metric space of the nilpotent approximation at point $q_0$ is exactly the Gromov–Hausdorff limit of the one parametric family of pointed metric spaces $(\lambda d_{SR}(\cdot, q_0))$ as $\lambda \to +\infty$.

To summarize, the book is a concise survey of the methods for motion planning of nonholonomic control systems by means of nilpotent approximation. It contains both the theoretical background and the explicit computational algorithms for solving this problem.

References


I. ZELENKO

TEXAS A&M UNIVERSITY

E-mail address: zelenko@math.tamu.edu