Nullstellensatz, Sheaf of Rings

an intuitive story

Daniel K. Rui - July 19, 2025

Abstract

I tried to make the "main story" accessible to even relatively fresh/green students (especially my "simplest/earliest/most natural substantive Examples"), but some examples/sidenotes/tangents may go too far afield. Ignore the examples/remarks for which you don't even recognize the words, but try to understand those for which you do.

The technical impetus for (affine) schemes really boils down to the following humble observation: I can always build "fractions" from a commutative unital ring (**cr1ng**) A (or thing *acted on by* A) — identical to the construction $\mathbb{Z} \rightsquigarrow \mathbb{Q}$ we are familiar with from childhood, for A an integral domain; and with a slight hiccup ("stabilization at sufficiently high powers") in the presence of zero divisors — and all those (subrings of) fractions assemble together in a "sheafy" way: the crucial sheaf *gluing property* comes (as in the subject of differential geometry or topology) from a "partition of unity", which one first sees in the commutative algebra setting in the form of Hilbert's Nullstellensatz. See the Nullstellensatz section if you just care about that; it is self-contained.

I also spent way too much time on the model theory stuff. That wasn't even meant to be a focus. And it's probably incomprehensible anyways. Alas...

Contents

1 Local-to-Global, and Sheaves

In analysis, we study many classes of functions from a (topological) space X to (typically) $\mathbb{R}$ or $\mathbb{C}$: continuous, smooth (C^k, C^∞), analytic/holomorphic, etc., where all of these notions also make sense on any open $U \subseteq X$. A very common thing to study is the relation between things happening locally and things happening globally. Going from global to local is *restriction* (not too interesting), but going from local to global one must *glue* (very interesting).

I tried to make the "main story" accessible to even relatively fresh/green students (especially my "simplest/earliest/most natural substantive Examples"), but some examples/sidenotes/tangents may go too far afield. Ignore the examples/remarks for which you don't even recognize the words, but try to understand those for which you do.

1.1 Local-to-global in various subjects, for context

For example, much effort is spent early on in one's study of topology/(graduate) real analysis on Urysohn's lemma or Tietze's extension theorem (which are tools that allow us to patch together local functions into a global function) and their consequences/applications: just off the top of my head we have Whitney embedding, density of continuous functions in various spaces like L^p, Riesz representation. See also the Hasse-Minkowski thereom in number theory; or Littlewood-Paley projections in harmonic analysis, or more generally the idea of dyadic decomposition in analysis for which Terry Tao has an amazing blogpost, from which I'll just sample one passage:

> "Very broadly speaking, one of the key advantages that dyadic models offer over non-dyadic models is that they do not have any "spillover" from one scale to the next. This spillover is introduced to us all the way back in primary school, when we learn about the algorithms for decimal notation arithmetic: long addition, long subtraction, long multiplication, and long division. In decimal notation, the notion of scale is given to us by powers of ten (with higher powers corresponding to coarse scales, and lower powers to fine scales), but in order to perform arithmetic properly in this notation, we must constantly "carry" digits from one scale to the next coarser scale, or conversely to "borrow" digits from one scale to the next finer one. These interactions between digits from adjacent scales (which in modern terminology, would be described as cocycles) make the arithmetic operations look rather complicated in decimal notation, although one can at least isolate the fine-scale behaviour from the coarse-scale digits (but not vice versa) through modular arithmetic. (To put it a bit more algebraically, the integers or real numbers can quotient out the coarse scales via normal subgroups (or ideals) such as $N \cdot \mathbb{Z}$, but do not have a corresponding normal subgroup or ideal to quotient out the fine scales.)
>
> It is thus natural to look for models of arithmetic in which this spillover is not present. One is first exposed to such models in high school, when the arithmetic of polynomials in one unknown t is introduced (i.e. one works with rings such as $\mathbb{Z}[t]$ or $\mathbb{R}[t]$ rather than $\mathbb{Z}$ or $\mathbb{R}$). For instance, to quotient one polynomial by another, one uses the polynomial long division (or synthetic division) algorithm, which is formally identical to long division for integers in decimal notation but without all the borrowing from one scale to the next. Here scales are represented by powers of t, rather than powers of 10. As with the reals or integers, the coarse scales can be contained in a normal subgroups (and ideals) such as $t^d \cdot \mathbb{R}[t]$, but now the fine scales can also be contained in normal subgroups (though not ideals) such as $\langle 1, t, \ldots, t^{d-1} \rangle$, the group generated by $1, t, \ldots, t^{d-1}$ (i.e. the group of polynomials of degree less than d). (From a category theory perspective, things are better here because various short exact sequences involving the scales are now split.)"

(Of course people have developed some methods to handle "spillover" between different scales = "local regions", e.g. induction on scales in harmonic analysis, cohomology to deal with cocycles, etc.) I should also mention that local-to-global is not the only method for building global functions with certain properties in analysis; it is also very common to take various limits (especially some kind of monotone limit) of other "worse" global functions to get a "good" global function — for example, in complex analysis, that is how we prove the Riemann mapping theorem, or use Perron's method to

build harmonic functions.

Or also Gromov opening his monograph "Sign and Geometric Meaning of Curvature":

> "The curvature tensor of a Riemannian manifold is a little monster of (multi)linear algebra whose full geometric meaning remains obscure. However, one can define using the curvature several significant classes of manifolds and then these can be studied in the spirit of the old-fashioned synthetic geometry with no appeal to the world of infinitesimals where curvature tensors reside. A similar interplay between infinitesimal quantities and visual features of geometric objects appears in all corners of geometry and analysis. The simplest example is provided by the equivalence of the two definitions of a monotone function
>
> $$\frac{df}{dt} \geqslant 0 \Longleftrightarrow f(t_1) \leqslant f(t_2) \text{ for } t_1 \leqslant t_2$$
>
> Then the infinitesimals of the second order bring along a geometrically more interesting phenomenon of convexity.
>
> $$\frac{d^2f}{dt^2} \geqslant 0 \Longleftrightarrow f\left(\frac{1}{2}(t_1+t_2)\right) \leqslant \frac{1}{2}(f(t_1+f(t_2))$$
>
> Our next example lies at the very verge of the Riemannian domain so we look at in a greater detail: The Second Fundamental Form and Convexity in the Euclidean space."

The above was a firehose of local-to-global ideas/principles for mathematical objects familiar to most students: namely essentially just functions on Euclidean space or manifolds (or slightly more generally, topological spaces sharing many properties with $\mathbb{R}^d$, e.g. spaces on which Urysohn's lemma holds), the real line (for e.g. dyadic decomposition), integers, etc.; but of course, the game becomes even more interesting (with less of the rougher "analytic"/"approximation"/"inequalities" flavor in my examples above, and more of an exact "algebraic"/"equalities" flavor) for more "general" notions of functions, like differential forms (de Rham cohomology) and sections of line bundles.

Gluing (& obstructions thereof) for differential forms, sections of bundles

In particular, it is with the sheaves of differential forms or sections of line bundles that one really starts to see how the topology/geometry of the space can affect the local-to-global gluing. If you didn't understand that preceding sentence, do not fear! — I will explain line bundles in this section (my remarks on differential forms are limited to 1 example in 1 paragraph, and half of that paragraph is thinking about the 1 example in terms of complex analysis).

For instance, the simplest substantive Example for the differential form story is how $\frac{1}{z}$ has an antiderivative locally (the complex integral $\int_{[z_0 \rightsquigarrow z]} \frac{1}{w}\, dw$ defines an antiderivative of $\frac{1}{z}$ on each disk $D(x_0, r)$ that does not contain 0), but these local antiderivatives do not glue together to get a global (smooth, or even continuous) antiderivative on $\mathbb{C} \smallsetminus \{0\}$. Equivalently phrased in real analysis, the differential form $\frac{y\,dx - x\,dy}{x^2+y^2}$ is closed (its derivative is 0) but not exact (is not the derivative $df = \partial_x f\, dx + \partial_y f\, dy$ of some function f). It is the topology of $\mathbb{C} \smallsetminus \{0\} \cong \mathbb{R}^2 \smallsetminus \{0\}$ that obstructs the gluing from local to global.

Actually, the above example "local antiderivatives of $\frac{1}{z}$ don't always glue to a global antiderivative" is also the simplest substantive Example for another phenomenon that often times one can find inverses of holomorphic functions locally, which don't glue to a global inverse. E.g. considering the holomorphic function $\exp : \mathbb{C} \to \mathbb{C} \smallsetminus \{0\}$, we know (e.g. by the general machinery of the lifting lemma) that for every simply connected open $U \subseteq \mathbb{C} \smallsetminus \{0\}$ (e.g. any open disk $\subseteq \mathbb{C} \smallsetminus \{0\}$), we can find some "pre-inverse" (right-inverse, or inverse you *pre*-compose with) $l : U \to \mathbb{C}$ s.t. $\exp \circ l = \mathrm{id}$. However, these "local inverses" (i.e. local branches of log) do not glue to a global holomorphic log function on $\mathbb{C} \smallsetminus \{0\}$. This

example is also in my bulleted list of examples of sheaves in §1.2.

The simplest substantive Example of a line bundle is the Möbius band. First, imagine (or better, physically hold) a rectangular strip of paper $[0, 2\pi] \times (-1, 1)$. Any (continuous, smooth) function $f : [0, 2\pi] \to (-1, 1)$ can be graphed as a (continuous, smooth) curve on this strip of paper. Now imagine (or physically do) bending the strip so that the ends meet, i.e. $\{0\} \times (-1, 1)$ becomes identified with $\{1\} \times (-1, 1)$, and we get a cylinder (see the picture here). Then, the graph of a continuous (resp. smooth) function $f : [0, 2\pi] \times (-1, 1)$ becomes a continuous (resp. smooth) curve on the cylinder iff $f(0) = f(2\pi)$ (resp. all one-sided derivatives match). We can think of these as graphs of functions $f : S^1 \to (-1, 1)$ (S^1 being the circle).

Alternatively, we could *twist* the strip before attaching/identifying the ends, i.e. $(0, y) \in \{0\} \times (-1, 1)$ becomes identified with $(2\pi, -y) \in \{2\pi\} \times (-1, 1)$ (see this MSE post for rigorous formulas). This produces a Möbius band, which I'll call M. There are a bajillion videos online if you need help visualizing (or, you could play with paper strips yourself). Then, the graph of a continuous (resp. smooth) function $f : [0, 2\pi] \times (-1, 1)$ becomes a continuous (resp. smooth) curve on M iff $f(0) = -f(2\pi)$ (resp. all one-sided derivatives are the negation of each other $f^{(k)}(0+) = -f^{(k)}(2\pi-)$). We can think of these as graphs of "functions on S^1 in M".

I illustrated one example of such a "function", in red on the picture on the below right. Notice that by the intermediate value theorem, in fact all "functions on S^1 in M" must have a zero (i.e. hit the "base" blue circle S^1). So somehow, the global geometry of M (the "twist") puts a very strong restriction on the behavior of "functions on S^1 in M": they must vanish somewhere! Note that the cylinder C has no such restriction. Note also that locally (on little "snippets" corresponding to some small $U \subsetneq S^1$), both C, M are identical to just a rectangle $U \times (-1, 1)$. In particular, locally, functions have no restriction on their behavior. This I think is the **fundamental/foundational observation behind line bundles**: locally, you can have whatever functions you want, but as you glue to get "functions" on a larger and larger domain, the global geometry then starts obstructing you.

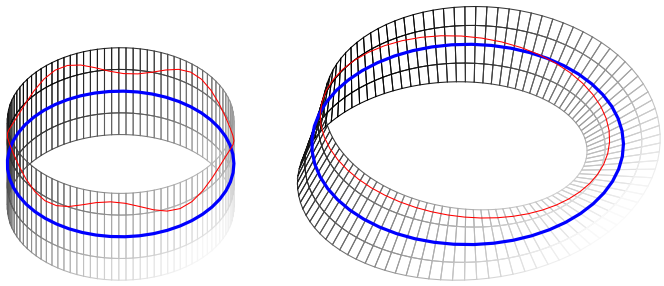
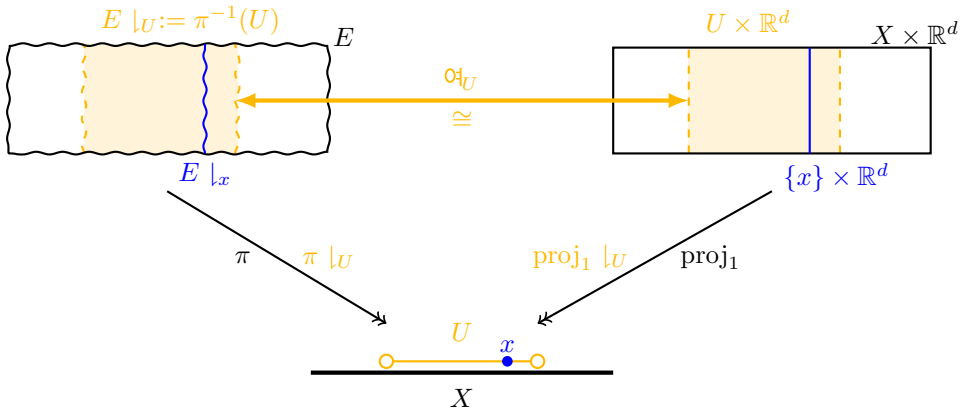


Figure 1: Line bundles on S^1 (graphics sourced mainly from this TEX.SE post)

The cylinder C and Möbius band M (in the graphic above, M, C are resp. the images of the two $[0, 2\pi] \times (-1, 1) \to \mathbb{R}^3$ maps $(x, l) \mapsto (\cos x, \sin x, 0) + \frac{1}{4} l \cdot (\cos(\frac{x}{2}) \cos x, \cos(\frac{x}{2}) \sin x, \sin(\frac{x}{2}))$ and $(x, l) \mapsto (\cos x, \sin x, l)$) are both examples of a line bundle on S^1:

- both consist of a "continuous family of lines" (line here meaning copies of $(-1, 1) \cong \mathbb{R}$), one for every $x \in X := S^1$ (namely ***the*** line ℓ_x going through x, a few of which you can see in the above graphics), where "continuous family" here means that any "snippet" of $L := C$ or M (also commonly

denoted with the letter E, maybe for "Euclidean") corresponding to a small open $U \subseteq X := S^1$ is identical to the Cartesian product $U \times (-1, 1)$ (just a "rectangle"). That was very handwavy, but we can formalize it as follows:



Definition 1.1: Line/Vector bundle

A continuous(/smooth/holomorphic) vector bundle E (over X) of rank d [for $d = 1$, called a line bundle] consists of the **Data** of

- a surjective continuous(/smooth/holomorphic) map

$$\pi : E \twoheadrightarrow X$$

  of topo. spaces(/smooth manifolds/complex manifolds) [$\rightsquigarrow$ notation $E \restriction_U := \pi^{-1}(U)$, etc.]

- **AND** a collection of homeomorphisms(/diffeomorphisms/biholomorphisms) [called "charts"; where $\{U_\alpha\}_\alpha$ must be an open cover of X]

$$\{\trivU_\alpha : E \restriction_{U_\alpha} \xleftarrow{\cong} U_\alpha \times \mathbb{R}^d\}_\alpha$$

  ($\mathbb{C}$ in place of $\mathbb{R}$ of course for complex, throughout this definition)

satisfying the **Property** that the above diagram commutes for every "U" $:= U_\alpha$.
In particular, this forces the blue fibers (above $x \in U \subseteq X$) to map homeomorphically(/...) to each other. In particular, $\trivU$ induces a $\mathbb{R}$-VS structure on the blue fiber $E \restriction_x$.

The **last Property** we ask for is that all $\trivU$ s.t. $U \ni x$ induce the *same* vector space structure on $E \restriction_x$, i.e. the vector addition function $+_x : E \restriction_x \times E \restriction_x \to E \restriction_x$ and scalar multiplication function $\cdot_x : \mathbb{R} \times E \restriction_x \to E \restriction_x$ (and choice of zero element $0_x : * \to E \restriction_x$), induced by different $\trivU$, are literally the same functions.
Equivalently, the *transition functions/"cocycles"* (for $x \in U_\alpha \cap U_\beta$) $[\trivU_\alpha \circ \trivU_\beta](x, \bullet) : \mathbb{R}^d \xRightarrow{\cong} \mathbb{R}^d$ are linear isomorphisms, i.e. $\transfU_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(d, \mathbb{R})$.
This is so that "sections" (defined below) have a well-defined addition and scaling law.

- Unpacking this definition in more "elementary" terminology, we have the **Data** of the "total space" $L := C$ or M

- and the **Data** of the "projection map" $\pi : L \twoheadrightarrow S^1$ (the map that crushes/projects/retracts every line ℓ_x to x),
- satisfying the **Property** that for every $x \in S^1$, there exists some open neighborhood $x \in U_x \subseteq S^1$ s.t. $\pi^{-1}(U_x)$ (the rigorous formulation of "the snippet of L corresponding to $U_x \subseteq S^1$") is homeomorphic to the "rectangle" $U_x \times \mathbb{R} \xleftarrow{\cong} \pi^{-1}(U_x) : \varphi_{U_x}$ (i.e. we want there to exist these "local trivializations" φ_{U_x}),
- where moreover φ_{U_x} restricted to each line $\ell_x := \pi^{-1}(\{x\}) \subseteq L$ is a homeomorphism $\ell_x \xrightarrow{\cong} \{x\} \times \mathbb{R}$ to the corresponding vertical line in the Cartesian product. This item formalizes the "family of lines" part, and the previous item formalizes the "continuous" part of "continuous family of lines".
- ... Satisfying one Last Property (please read this AFTER reading below on sections): different local trivializations $\varphi_{U_1}, \varphi_{U_2}$ coming from 2 sufficiently small opens $U_1, U_2 \ni x$ give 2 different homeomorphisms to the line ℓ_x:

$$\{x\} \times \mathbb{R} \xleftarrow{\varphi_{U_1}}_{\cong} \ell_x \xrightarrow{\varphi_{U_2}}_{\cong} \{x\} \times \mathbb{R}.$$

  We ask that the composition $\varphi_{U_2} \circ \varphi_{U_1}^{-1}$ is an isomorphism of $\mathbb{R}$-vector-spaces. The purpose of this is to make addition (and scaling) of sections well defined, so that the set of sections on any (open) domain $D \subseteq S^1$ forms an $\mathbb{R}$-VS. It is a good exercise to work through the details yourself, to see why exactly this extra property is needed to make the addition (and scaling) well-defined. One may ask: can we multiply functions? Unfortunately no. Try to think through this in the case of the Möbius band. The issue is that you can't find a cts. choice for the multiplicative unit 1, since as I discuss below all cts. sections must cross zero! Alternatively, going back to fns. on $[0, 2\pi]$ satisfying $f(0) = -f(2\pi)$, observe that $f_{1,2}(0) = -f_{1,2}(2\pi) \implies f_1 \cdot f_2(0) = +f_1 \cdot f_2(2\pi)$! But things work well for addition and scaling. It is very interesting that it is fairly easy to allow addition for sections of line bundles, but not multiplication.
- *Last remarks:* notice that the local trivializations φ_{U_x} are not part of the **Data**, but rather their existence is one of the **Properties**. This is in contrast to smooth manifolds, where the choice of chart/atlas is part of the data. This is because a different chart/atlas (= "smooth structure") on a manifold gives rise to different sets of smooth functions; but here, different collections of local trivs. satisfying the above properties will give rise to the same sections.

Honestly, Wiki or this PDF explains this better/cleaner than me. Wiki defines fiber bundles for general fibers F (instead of just $(-1, 1) \cong \mathbb{R}$).

- A (continuous, smooth, etc.) "function on the base S^1 to $\mathbb{R}$ in the total space $L := C$ or M" (the correct language is "global section of the line bundle L") is a function $s : S^1 \to L$ s.t. every point $x \in S^1$ is sent to another point on "its line" ℓ_x, s.t. locally on $U_x \ni x$, translating through the homeomorphism φ_{U_x}, the image $s(U_x) \subseteq \pi^{-1}(U_x) \subseteq L$ looks like the graph of a (continuous, smooth, etc.) function $U_x \to (-1, 1) \cong \mathbb{R}$.

- *Last remarks:* the cylinder C is called the trivial line bundle on S^1 because not only does it locally look like the Cartesian product $S^1 \times \mathbb{R}$, it *globally* is exactly $S^1 \times \mathbb{R}$.

  Also, everything above goes through with $\mathbb{R}^d$ in place of $\mathbb{R} \cong (-1, 1)$; those are called vector bundles, and sections of those are intuitively "functions on the base X to $\mathbb{R}^d$ in the total space L" or "vector-valued functions on the base X in the total space L".

  Also, the tangent bundle for a smooth manifold is a foundational example of a vector bundle (in which the Last Property above arises naturally); maybe it sounds scary, but if you just look on

coordinate patches, it's actually quite easy — indeed, "there is no calculus on manifolds, but only calculus on $\mathbb{R}^d$ and pushforwards and pullbacks".

In other words, a line bundle on a topo. space X is the **Data** of a space L **AND** a map $\pi : L \twoheadrightarrow X$, satisfying the **Property** that locally, $L \xrightarrow{\pi} X$ looks like the Cartesian product with the $\mathbb{R}$-VS $\mathbb{R}$ and (1st coordinate) projection $(X \times \mathbb{R}) \xrightarrow{\text{proj}_1} X$ (yes, this is vague/loose; the rigor is above. Or on Wiki).

And a "function on X in L" (the correct language is "global section of the line bundle L") is a map $s : X \to L$ s.t. $s(x) \in \pi^{-1}(\{x\}) \subseteq L$, and which locally (translated through the relevant homeomorphism) looks like the graph of a (continuous, smooth, etc.) function $U_x \to \mathbb{R}$. Perhaps one can think of sections of a given l.b. $L \twoheadrightarrow X$ as functions $X \to \mathbb{R}$ that are continuous except at certain places, and can only be discontinuous there in some extremely constrained way, like how sections of the Möbius bundle $\leftrightarrow$ functions on S^1 continuous except at one point p s.t. $\lim_{x \to p^+} f(x) = -\lim_{x \to p^-} f(x)$.

The formalism is, unfortunately, unavoidably, really just kind of ugly, and is difficult to write out fully and still intuitively (I doubt I succeeded). The intuition is quite simple though: a line bundle L is a "continuous family of lines on/over X" that "locally looks like the Cartesian product $(X \times \mathbb{R}) \xrightarrow{\text{proj}_1} X$", and a "function on X in L" (e.g. "function on S^1 in cylinder C" or "function on S^1 in Möbius band M") is exactly what you think it is, and for some reason it just takes a lot of symbols to write down what exactly you think it is. I swear it's a lot easier to explain/understand if I get to draw and wave my arms in front of you. O, the tragedy of the written word...

1.2 Brief Introduction to Sheaves

At this point, I have hopefully convinced you that this local-to-global business is very interesting. Especially in the last 2 examples, where we had lots of "functions" (differential forms, or sections of line bundles) locally, but somehow encountered obstructions to gluing those local "functions" together to get a global "function". A sheaf is just a collection of all the data of local functions, and how they restrict to one another (i.e. the global-to-local half of the business, which I said at the very beginning was the uninteresting half), satisfying a few extra (very intuitive) "sanity check" properties.

- The sheaf of ($\mathbb{R}$-valued) continuous ("cts.") functions $C_X(-)$ on a topo. space X consists of the (presheaf) **Data** of a map $[U \mapsto C_X(U)]$ taking any open $U \subseteq X$ to $C_X(U) :=$ the set (or actually $\mathbb{R}$-VS, or actually $\mathbb{R}$-Alg.) of cts. functions $U \to \mathbb{R}$, **AND** restriction maps $|^U_V : C_X(U) \to C_X(V)$ that relate functions f on larger opens U to functions $g = f|^U_V$ on smaller opens $V \subseteq U$.

  Remark: this **Data** is the same data as a functor $\mathsf{C}_X : \underline{\textbf{Opens}}_X \to \underline{\textbf{Set}}$ (or actually, $\to \underline{\mathbb{R}\textbf{-Alg}}$)

  The **Data** above furthermore satisfies the **Properties** (which are all completely intuitive in the C_X example) of: (presheaf) restriction identity and compatibility (the same properties as the identity + composition properties for the functor C_X); and (sheaf) Locality/Uniqueness and Gluing. The precise statements are written well on Wikipedia. See also this exposition by Agrios (2022).

- Along with $C_X(-)$, we have also basically any variation we can think of (let me omit the subscript X): $C_{\text{b}}(-)$ (bounded cts.), $C_0(-)$ (cts. functions $\to 0$ at ∞, in the sense of Alexandroff 1-point cpct.), $C^{\text{lsc}}(-)$ (lower semi-cts.), $C^\times(-)$ (cts. no-where vanishing = invertible), and so on.

  There is one major exception to this parade: $C_{\text{c}}(-)$ (compactly supported continuous functions) do NOT form a sheaf, since restricting a function compactly supported in U to $V \subseteq U$ may make it not compactly supported in V. Instead, $C_{\text{c}}(-)$ forms a cosheaf (restriction works in the opposite direction $V \supseteq U$), whose dual is a sheaf, namely the sheaf of distributions (see Wiki).

- On a (smooth) manifold M (e.g. concretely $M :=$ open $U \subseteq \mathbb{R}^d$, or S^1, or the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$), smooth functions $C^k_M(-)$ or $C^\infty_M(-)$ (with the obvious $\restriction$ maps) form sheaves of $\mathbb{R}$-algebras.

  Also $\Omega^d_M(-)$ the differential n-forms on M form a sheaf of $\mathbb{R}$-VS (with the obv. restr. maps).

- On $M := \mathbb{C}^d$ (or more generally, a complex manifold M, e.g. Riemann surface, e.g. again a torus $\mathbb{C}/\Lambda$ for any lattice $\Lambda = \omega_1\mathbb{Z} + \omega_2\mathbb{Z}$), holomorphic/analytic functions $\mathcal{O}^{\text{hol/an}}_M(-) := \mathscr{Hol}_M(-)$ (with the obvious restriction maps) forms a sheaf of $\mathbb{C}$-algebras.

  On $\mathbb{C}^d$, algebraic/rational functions with nowhere-zero denominators form a sheaf of $\mathbb{C}$-algebras $\mathcal{O}^{\text{alg}}_{\mathbb{C}^d}(U) := \{\frac{f}{g} : f, g \in \mathbb{C}[x_1, \dots, x_d], g \text{ never} = 0 \text{ in } U\}$ (with the obv. restr. maps). This example will be central. However, it should NOT be obvious at this point that $\mathcal{O}^{\text{alg}}_{\mathbb{C}^d}$ satisfies Gluing.

- Actually, the simplest substantive Example from my paragraph on differential forms, "local antiderivatives of $\frac{1}{z}$", is also the simplest substantive Example for a more general construction: the sheaf of inverses of a holomorphic function — see this MSE comment. Indeed, consider the holomorphic function $\exp : \mathbb{C} \to \mathbb{C} \smallsetminus \{0\}$. Let the "base space" X be $X := \mathbb{C} \smallsetminus \{0\}$ (with the standard Euclidean topology). Then on X we can define a sheaf of sets by $\mathcal{O}(U) := \{f : U \to \mathbb{C} : \exp \circ f = \text{id}\} \subseteq \mathscr{Hol}(U)$ (with the obvious restriction maps).

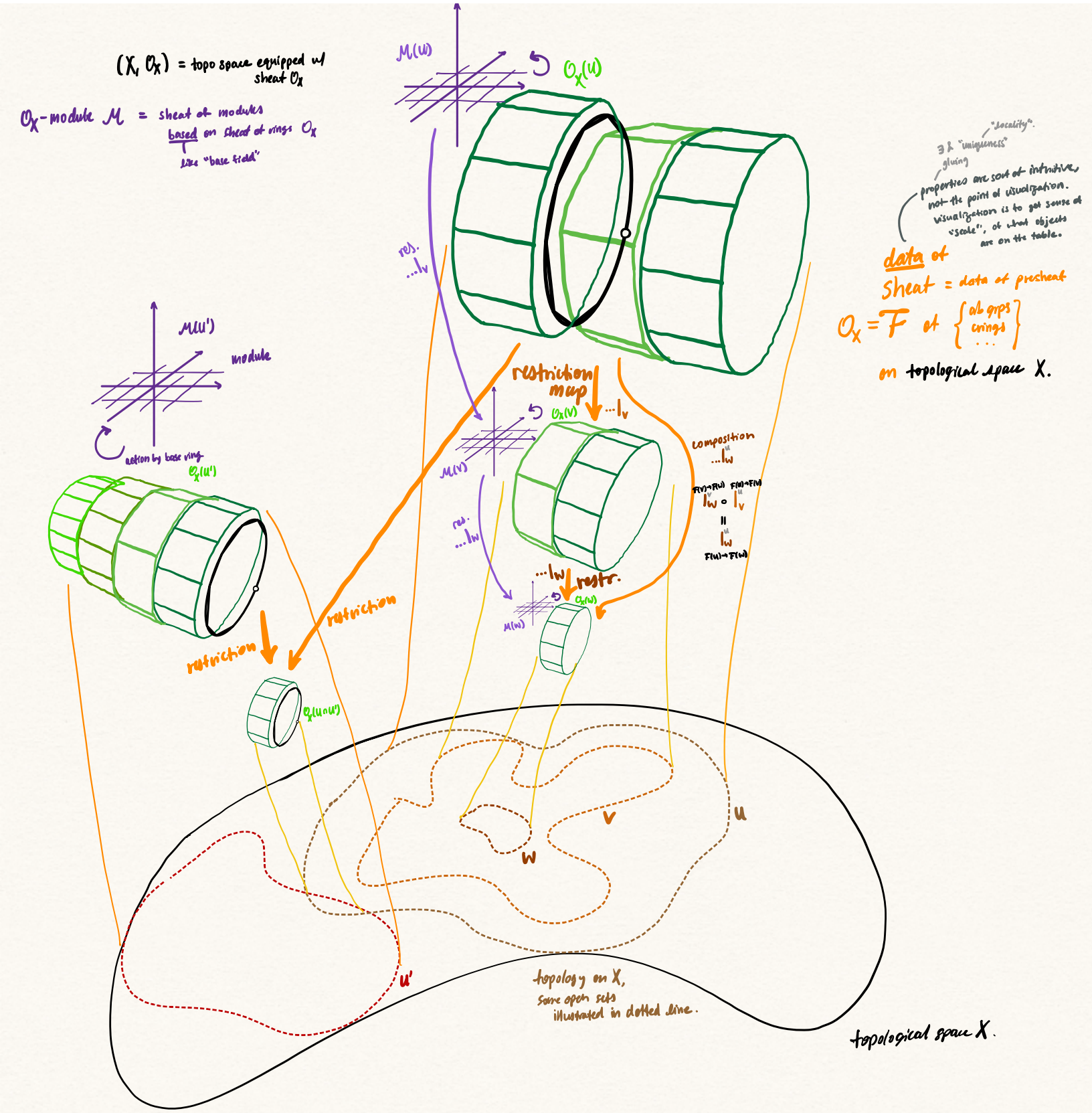
  If one wants to be fancy, observe that because $e^{2\pi i} = 1$, $f \in \mathcal{O}(U) \implies f + 2\pi i k \in \mathcal{O}(U)$ for all $k \in \mathbb{Z}$, the above sheaf is actually a sheaf of $\mathbb{Z}$-sets.

- For a line bundle $L \overset{\pi}{\twoheadrightarrow} X$, the (local) sections $U \mapsto \Gamma(U, L) =: H^0(U, L)$ (with the obvious restriction maps) form a sheaf of $\mathbb{R}$-vector-spaces.

As you/After you read the definition of sheaf 10 times (seriously, you should. On 10 separate days.) and check the properties for the above examples, the following picture I produced may or may not be useful:

Below, I illustrate a sheaf $\mathcal{O}_X(-)$ of cr1ngs on a topo. space X (open sets are brown-ish colored), and a (purple-ish colored) $\mathcal{O}_X$-module $\mathscr{M}(-)$ ("sheaf of modules based on the sheaf of cr1ngs $\mathcal{O}_X$", where for every open $U \subseteq X$, we get a module $\mathscr{M}(U)$ acted on by the cr1ng $\mathcal{O}_X(U)$).

I depict cr1ngs (green-ish colored) in the same way that this YT video depicts f.g. abelian groups $\mathbb{Z}^r \times \mathbb{Z}/t_1\mathbb{Z} \times \dots \mathbb{Z}/t_n\mathbb{Z}$ (which happen to also be cr1ngs, on top of f.g. ab. grps.). The restriction maps are in orange-ish colors.


$(X, \mathcal{O}_X)$ = topo space equipped w/ sheaf $\mathcal{O}_X$
$\mathcal{O}_X$-module $\mathcal{M}$ = sheaf of modules based on sheaf of rings $\mathcal{O}_X$
like "base field"
$\mathcal{M}(U)$
$\mathcal{O}_X(U)$
"locality". ∃ & "uniqueness" gluing
properties are sort of intuitive, not the point of visualization. visualization is to get sense of "scale", of what objects are on the table.
data of sheaf = data of precheaf
$\mathcal{O}_X = \mathcal{F}$ of {ab grps, crings, ...}
on topological space X.
res. $\dots|_V$
restriction map
$\dots|_V$
$\mathcal{M}(U')$
module
action by base ring
$\mathcal{O}_X(U')$
$\mathcal{O}_X(V)$
$\mathcal{M}(V)$
composition $\dots|^U_W$
$\dots|^U_W \circ \dots|^U_V = \dots|^U_W$
res. $\dots|_W$
$\dots|_W$ restr.
$\mathcal{O}_X(W)$
$\mathcal{M}(W)$
restriction
restriction
$\mathcal{O}_X(U \cap U')$
U
V
W
U'
topology on X, some open sets illustrated in dotted line.
topological space X.

2 The Sheaf for $\mathbb{C}[\vec{x}]$ (and then for all A)

Prereq: you should probably know the def. of a sheaf before reading this section (see my attempted intro above; or Wiki or e.g. Agrios2022). It turns out that the sheaf (that we don't know satisfies Gluing yet) of rational functions (already introduced above) $\mathcal{O}^{\text{alg}}_{\mathbb{C}^d}(U) := \{\frac{f}{g} : f, g \in \mathbb{C}[x_1, \ldots, x_d], g \text{ never} = 0 \text{ in } U\}$ (with obv. $\restriction$) is the most natural example of the sheaf of rings construction to start with. And amazingly, its proof [I later found MSE] will contain within it "all the germs of generality" (to quote Hilbert).

Facts about $\mathbb{C}$ (or ACF) polynomials $A = \mathbb{C}[\vec{x}]$: each $f \in A$ vanishes on $V(f \mid \mathbb{C}^d) \subseteq \mathbb{C}^d$, and $\frac{1}{f}$ is a well-defined function on the (Euclidean-topology-)open set $\mathbb{C}^d \smallsetminus V(f \mid \mathbb{C}^d) =: D(f \mid \mathbb{C}^d) \subseteq \mathbb{C}^d$. The collection of $D(f)$ is closed under finite intersection. The following theorem tells us what kinds of rational functions are well-defined functions on the "distinguished open" $D(f)$:

Theorem 2.1: Hilbert Nullstellensatz over $\mathbb{C}$ (proven in another section)

IF g_i, f are s.t. "the g_i force f to vanish": [$g_1, \ldots, g_n$ vanish at $\vec{x}_0 \in \mathbb{C}^d \implies f$ vanishes at $\vec{x}_0$] — or equivalently, $\bigcap_{i=1}^n V(g_i \mid \mathbb{C}^d) \subseteq V(f \mid \mathbb{C}^d) \iff D(f \mid \mathbb{C}^d) \subseteq \bigcup_{i=1}^n D(g_i \mid \mathbb{C}^d)$,

THEN the only possible explanation is an algebraic relation: $f^r = \sum_{i=1}^n a_i g_i$ for some $a_i \in A$ (*"the ideal $(g_1, \ldots, g_n)$ 'stably contains' f"*) — or equivalently, $g_1, \ldots, g_n$ generate the unit ideal in $A[\frac{1}{f}] \iff$ we have a "partition of unity" $1 = \sum_{i=1}^n r_i g_i$ for rational functions $r_i \in A[\frac{1}{f}]$.

(Of course, conversely, such an algebraic relation implies the forcing of vanishing.)

In particular,

Corollary 2.1: Functions on distinguished open

The only well-defined rational functions on $D(f \mid \mathbb{C}^d)$ are $R[\frac{1}{f}]$, because given $\frac{p}{q} \in \mathbb{C}(\vec{x})$ s.t. $V(q) \subseteq V(f)$, then the above theorem tells me

$$f^r = aq \implies \frac{1}{q} = \frac{a}{f^r} \implies \frac{p}{q} \in A\Big[\frac{1}{f}\Big].$$

This also implies

Proposition 2.1: Restriction in sheaf of rings

IF $D(f) \subseteq D(g) \iff V(g) \subseteq V(f)$ (which recall above is $\iff f^r = ag$),

THEN any s on $D(g)$ (i.e. $s \in A[\frac{1}{g}]$) restricts to $\tilde{s}$ on $D(f)$ (i.e. $\tilde{s} \in A[\frac{1}{f}]$ s.t. $s = \tilde{s}$ on $D(f)$, i.e. as elements of $A[\frac{1}{f}]$).

Conversely (this works in any cr1ng A!), if every section in $A[\frac{1}{g}]$ restricts to one in $A[\frac{1}{f}]$, i.e. IF for every $s \in A[\frac{1}{g}]$ there is $\tilde{s} \in A[\frac{1}{f}]$ s.t. $s = \tilde{s}$ as elements of ... actually $A[\frac{1}{fg}]$, since elts. of $A[\frac{1}{f}]$ and $A[\frac{1}{g}]$ both can be interpreted there, since the universal property lifts the "over 1" map $\text{무} := [a \mapsto \frac{a}{1}] : A \to A[\frac{1}{fg}]$ to a unique $A[\frac{1}{f}] \to A[\frac{1}{fg}]$),

THEN in fact $f^r = ag$.

As I mentioned above, the most natural/intuitively obvious place to compare (i.e. check equality) for rational functions in $A[\frac{1}{f}]$ and $A[\frac{1}{g}]$ is $A[\frac{1}{fg}]$ (it's exactly the rigorous semantic context in which people perform the familiar syntactic fraction manipulations). Just like how in sheaf theory one checks compatibility for different sections $s_{1,2}$ on opens $U_{1,2}$ on $U_1 \cap U_2$. Thus, we *should* expect $D(f) \cap D(g) = D(fg)$ for all $f, g \in A$. This is obviously true for $A = \mathbb{C}[\vec{x}]$, but this sheaf theoretic thinking is very general, and provides a *prediction* for the abstract theory we are about to develop for any cr1ng A.

Recall the Gluing Property of schemes says that assuming $D(f) \subseteq \bigcup_{i=1}^n D(g_i) \implies D(f) = \bigcup_{i=1}^n D(fg_i)$ (in what follows, I'm confident that taking $s_i \in A[\frac{1}{g_i}]$ instead of $s_i \in A[\frac{1}{fg_i}]$ will produce exactly the same results),

- IF we have $s_i \in A[\frac{1}{fg_i}]$ s.t. every s_i, s_j are equal in $A[\frac{1}{fg_ig_j}]$ — which recall means equality $s_i(fg_ig_j)^e = s_j(fg_ig_j)^e$ as elements in A for some (and hence all) large $e \geqslant 1$ (for large enough e, we can cancel the denominators in s_i, s_j to get actual elements in A, which is what I meant by "equality as elements of A"). *("s_i equals s_j 'stably relative to' fg_ig_j")*
- THEN there should be some $s \in A[\frac{1}{f}]$ s.t. for every i, $s = s_i$ in $A[\frac{1}{fg_i}]$.

Or in looser/less rigorous phrasing, "$A[\frac{1}{fg_1}] \cap \ldots \cap A[\frac{1}{fg_n}] \subseteq A[\frac{1}{f}]$".

Note also that this Gluing, which I'll say "corresponds to the covering $D(f; A) \subseteq \bigcup_i D(g_i; A)$", is equiv. to the Gluing which "corresponds to the covering $D(1; A[\frac{1}{f}]) \subseteq \bigcup_i D(g_i; A[\frac{1}{f}])$". This already doesn't make sense (yet) for $A = \mathbb{C}[\vec{x}]$, but as I said above, this sheaf theoretic thinking (this Gluing of the rings of fractions = "functions on different opens") is very general, and provides another *prediction* for the abstract theory we are about to develop for any cr1ng A.

The proof that Gluing Property holds is as follows:

- Fix $s_i \in A[\frac{1}{fg_i}]$, i.e. some fraction-looking expression $\frac{\bullet}{(fg_i)^\bullet}$, where $s_i = s_j$ in $A[\frac{1}{fg_ig_j}]$, i.e.
- Let M be a bigger integer than any power g_i^e that appears in the denominator of any of the $s_1, \ldots, s_n$.
- Use the covering assumption, which is equivalent to the "generating unit ideal of $A[\frac{1}{f}]$" property above to get a "partition of unity"
$$1 = \sum_{i=1}^n r_i g_i^M$$
for some $r_i \in A[\frac{1}{f}]$
- Then consider $s := \sum_{i=1}^n s_i r_i g_i^M$ (the intuition is that if all the s_i were all "literally/syntactically equal" to s, then $s = s \cdot 1 = s \cdot \sum_{i=1}^n r_i g_i^M$). The power g_i^M is chosen large enough to cancel any power of g_i in the denominator of s_i, so indeed all these fractions live in $R[\frac{1}{f}]$.
- One can check all desired/necessary equalities. Indeed, to show $s = s_1$ in $R[\frac{1}{fg_1}]$, observe that
$$\begin{aligned}(fg_1)^M \cdot s := (fg_1)^M \cdot \sum_{i=1}^n s_i r_i g_i^M &= \sum_{i=1}^n r_i \cdot s_i (fg_1g_i)^M \\ &= \sum_{i=1}^n r_i \cdot s_1 (fg_1g_i)^M = (fg_1)^M \cdot \sum_{i=1}^n s_1 r_i g_i^M = (fg_1)^M \cdot s_1 \cdot 1\end{aligned}$$

We have thus proven:

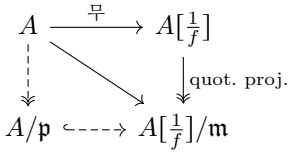
Theorem 2.2: The Sheaf for $\mathbb{C}[\vec{x}]$

For $A = \mathbb{C}[\vec{x}]$, forming localizations $A[\frac{1}{f}]$ for $\forall f \in \mathbb{C}[\vec{x}]$, we get a sheaf $\mathcal{O}(D(f \mid \mathbb{C}^d)) := A[\frac{1}{f}]$ on the (finite-intersection-closed) base of "Distinguished Opens" ...

... where in particular the Gluing Property (given a cover) for sheaves was proven just from the "generating unit ideal"/"Nullstellensatz" property that covers satisfy (in fact are equivalent to) by the Nullstellensatz.

2.1 Open covers in general cr1ngs A

There is one last piece of the puzzle, from commutative algebra: a common "reflex"/"instinct" one develops in commutative algebra (at the cost of using the Axiom of Choice) is to recognize that if something does not generate the unit ideal, then it is contained in some maximal ideal $\mathfrak{m}$. And that pulling back (i.e. preimage) a maximal ideal $\mathfrak{m} \subsetneq A[\frac{1}{f}]$ through the "over 1" map $\fallingdotseq := [a \mapsto \frac{a}{1}] : A \to A[\frac{1}{f}]$, we get a prime ideal $\mathfrak{p}$ in A. One can see this from the field/integral domain characterization of maximal/prime ideals: the 1st isomorphism theorem factors every morphism into an $\twoheadrightarrow$ (a quotient projection to the quotient by the kernel) followed by an inclusion $\hookrightarrow$



where $A/\mathfrak{p} \hookrightarrow$ a field means it must be an integral domain, $\iff \mathfrak{p}$ prime. Note in particular that $f \notin \mathfrak{p}$, because if it were, then $\fallingdotseq(f) \in \mathfrak{m}$, but f is a unit in $A[\frac{1}{f}]$ and maximal ideals can't contain units.

So, we have shown that $[g_1, \ldots, g_n$ generate the unit ideal in $A[\frac{1}{f}]] \iff [g_1, \ldots, g_n$ is contained in a maximal ideal $\mathfrak{m} \subsetneq A[\frac{1}{f}]] \implies$ [there is a prime ideal $\mathfrak{p} \subsetneq A$ s.t. $g_1, \ldots, g_n \in \mathfrak{p}$ but $f \notin \mathfrak{p}$]. The last implication's converse ($\impliedby$) is also true, and left as an exercise to the reader.

The takeaway is that the "generating unit ideal"/"Nullstellensatz" property (w.r.t. fixed f and $g_1, \ldots, g_n$) is $\iff$ [$\forall \mathfrak{p}$ prime ideal, $g_1, \ldots, g_n \in \mathfrak{p} \implies f \in \mathfrak{p}$], which can be rephrased as the set-theoretic covering expression $D(f; A) \subseteq \bigcup_{i=1}^n D(g; A) \iff \bigcap_{i=1}^n V(g_i; A) \subseteq V(f; A)$, **as long as we define** $V(h; A)$ to be the set of prime ideals $\mathfrak{p}$ containing h, or equivalently $D(h; A) := \{\mathfrak{p} \lhd A : \mathfrak{p} \not\ni h\}$!!!

Key Point 2.1

The remarkable fact is that everything above goes through, practically verbatim, as long as instead of considering opens $D(f \mid \mathbb{C}^d)$ consisting of points in $\mathbb{C}^d$, we consider $D(f; A)$ consisting of "points" that are prime ideals $\mathfrak{p}$ of A.
In fact, above is a *complete construction* of the sheaf of rings of any cr1ng A, but an *incomplete account* for $A = \mathbb{C}[\vec{x}]$, since I haven't proven the Nullstellensatz yet! You see the "meat" of the Nullstellensatz is to go between (i.e. prove equivalent) the "complex pts" and "prime ideal pts" perspective on covers $D(f \mid \mathbb{C}^d) \subseteq \bigcup_{i=1}^n D(g_i \mid \mathbb{C}^d) \iff D(f; \mathbb{C}[\vec{x}]) \subseteq \bigcup_{i=1}^n D(g_i; \mathbb{C}[\vec{x}])$.

Also, we haven't constructed the full sheaf $\mathcal{O}^{\text{alg}}_{\mathbb{C}^d}(-)$, since we currently only have it for the open sets $D(f \mid \mathbb{C}^d) \subseteq \mathbb{C}^d$. But for general U, we can get $\mathcal{O}^{\text{alg}}_{\mathbb{C}^d}(U)$ as a colimit over $D(f \mid \mathbb{C}^d) \supseteq U$. Colimits are very important (though like basically everything else I've raced through in this document, they are kind of technical and it takes a while to get used to them), so take the time now to really understand them in this rather concrete setting.

3 Philosophy and Complaints

Facts from commutative algebra: given cr1ng A, we have localizations $A[\frac{1}{f}]$, and for $\mathfrak{p}$ a prime ideal, $A \setminus \mathfrak{p}$ is multiplicatively closed so can form localization $A_{\mathfrak{p}}$.

I think of the sheaf (so data $A, A[\frac{1}{f}], A_{\mathfrak{p}}$, etc.) as *real*, and the distinguished opens as *kind of fake*, and "points" = prime ideals making up those distinguished opens as *really fake*. As you saw from the above construction, we started with the *real* data ("the castle floating in the sky"), and generated a foundation/base on which this castle can now rigorously stand. One support for my above ontology (the sheaf of rings is real but the space is fake) is this quote from MO

> "This means that the usual definition of Spec A with prime ideals is *a* correct one (as long as one is working in classical logic with the axiom of choice) but it does not mean that it is the only correct definition: You can, for example, replace Spec by any other topological space or, more generally, by any other site such that the sheaf topos over it is still equivalent to X."

One common thing people bring up when trying to motivate prime ideals as points is first look at ring homomorphisms $A \to K$ for fields K: see this MSE answer, and this MO answer from which I quote

> "A locally-ringed space is a topological space with a sheaf of rings ("the sheaf of (admissible) functions on the space") such that the stalks are local rings. Why should the stalks be local rings? Because even if you generalize (or specialize) your notion of a function, you want to have the notion of a function vanishing at a point, and those functions that vanish at a point should be a very special (read: unique maximal) ideal in the stalk. Alternatively, the values of functions at points should be elements of fields; if the value is an element of some other kind of ring, then you're not really looking at a point."

This MO link further discusses the equivalence between the "homs to fields" perspective and prime ideals perspective.

My problem with this is that this fact/wish that stalks are local rings, and that values of functions at points should be fields, are good *a postieri* rationalizations that the theory we've developed has nice properties. But (at least in my eyes, as I was trying to learn this subject), they were not really compelling enough *a priori* to justify launching this entire campaign to first define points of a space as prime ideals, then opens, then the sheaf.

Furthermore, focusing on homs to fields only works for the points (of the underlying space), not the sheaf on the space (which again, is the star of the show), as discussed in the footnote of pg. 12 of this thesis of Ingo Blechschmidt.

4 Nullstellensatz

Let us now return to the last step in §2. I will first present it taking one model theoretic result as a black box (since ultimately I think it is a good idea to think of the Nullstellensatz conceptually as a model theoretic result, instead of an commutative algebra/algebraic geometry result), and then I will make a deviation in that outline in order to give a complete and self contained proof here.

In slightly more detail, in the model theory route, we will at one point encounter a field L (that is built from $\mathbb{C}[\vec{x}]$ or $\mathbb{C}[\vec{x}][\frac{1}{f}]$ by quotienting by a maximal ideal) that has a certain Nice Property (constructed so that some polynomials have roots and some don't), which is inherited by the algebraic closure $\overline{L}$. By the completeness of the theory ACF_0, every algebraically closed char. 0 field satisfy the same 1st order sentences, so in fact $\mathbb{C}$ also has the Nice Property.

This 9-page REU paper by Ford (2007) introduces model theory (from the very beginning) and follows essentially this route. The 13th lecture in a video series by Artem Chernikov (for the model theory quarter of UCLA's mathematical logic sequence 220ABC) teaches this result, with a full background of model theory — I highly recommend it. Take this as an advertisement to learn some model theory! More advertising: https://mathoverflow.net/questions/9667/what-are-some-results-in-mathematics-that-have-snappy-proofs-using-model-theory.

The deviation will be to prove Zariski's [f.g. field algebraicity] lemma (from this 3-page note by Allcock: "This is the simplest proof of the Nullstellensatz that I have been able to come up with."; also "one-line undergraduate proof" from Azarang), that shows that in fact because L is a field extension of $k := \mathbb{C}$ that is finitely generated as a k-algebra, it must be an algebraic field extension of k. But $k := \mathbb{C}$ is algebraically closed, so L must in fact be $\mathbb{C}$ itself.

This sidesteps the need to use the completeness of the theory ACF_0, as in the model theory proof (we used it there to transfer a Nice Property from $\overline{L}$ to $\mathbb{C}$; but here we show actually $\overline{L} = L = \mathbb{C}$!). The lemma is easy over $\mathbb{C}$ with a bit of complex analysis background, and there is a slightly more clunky algebraic proof that works for all fields k.

Remark: Allcock's proof of Zariski's lemma is somewhat similar in spirit to this short proof on MO for the special case of uncountable algebraically closed fields, and another MO answer (by Brian Conrad) linked therein that pushes the special case to the general case.

For Nullstellensatz-like results for polynomials not over an algebraically closed field, see https://mathoverflow.net/questions/120568/hilberts-nullstellensatz-on-polynomials-with-integer-coefficients, and the quite thorough https://kconrad.math.uconn.edu/blurbs/ringtheory/maxideal-polyring.pdf

For other proofs of the Nullstellensatz (I personally did not read these, since they are just too long...), here are links to a proof by Munshi https://www.math.uchicago.edu/~may/PAPERS/MunshiFinal2.pdf, and an algorithmic proof by Tao https://terrytao.wordpress.com/2007/11/26/hilberts-nullstellensatz/ using elimination theory (amusingly using both forward and backward induction crucially). Tao has a nice quote:

> "Like many results of the "The only obstructions are the obvious obstructions" type, the power of the nullstellensatz lies in the ability to take a hypothesis about non-existence (in this case, non-existence of solutions to $P_1(x) = \ldots = P_m(x) = 0$) and deduce a conclusion about existence (in this case, existence of $Q_1, \ldots, Q_m$ such that $P_1Q_1 + \ldots + P_mQ_m = 1$). The ability to get "something from nothing" is clearly going to be both non-trivial and useful."

4.1 Beginning proof of Nullstellensatz

Recall the statement of the Nullstellensatz from before. I will state and prove the contrapositive.

Theorem 4.1: Hilbert Nullstellensatz over $\mathbb{C}$

IF $g_1, \dots, g_n$ do NOT generate the unit ideal in $A[\frac{1}{f}]$ — or equiv., $g_1, \dots, g_n \in \mathfrak{m}$ a maximal ideal of $A[\frac{1}{f}]$;

THEN $\bigcap_{i=1}^n V(g_i \mid \mathbb{C}^d) \not\subseteq V(f \mid \mathbb{C}^d)$ — or equiv., $\exists$ a solution $\overline{a} \in \mathbb{C}^d$ to the system $\left\{ \begin{smallmatrix} \text{all } g_i(\overline{a})=0 \\ f(\overline{a}) \neq 0 \end{smallmatrix} \right.$.

(Of course, conversely, such solution to the system implies the impossibility of $g_1, \dots, g_n$ generating the unit ideal in $A[\frac{1}{f}]$.)

First half of proof of Nullstellensatz: of course, given $g_1, \dots, g_n \in \mathfrak{m}$ a maximal ideal of $A[\frac{1}{f}]$, we quotient by $\mathfrak{m}$ to get a field

$$L \coloneqq \mathbb{C}[x_1, \dots, x_d]\big[\tfrac{1}{f}\big]/_{\mathfrak{m}}.$$

Observe that in this field, we have a solution to the system $\left\{ \begin{smallmatrix} \text{all } g_i(\overline{a})=0 \\ f(\overline{a}) \neq 0 \end{smallmatrix} \right.$, namely the cosets $\frac{x_1}{1} + \mathfrak{m}, \dots, \frac{x_d}{1} + \mathfrak{m} \in L$, because g_i evaluated at those land in $\mathfrak{m}$ (i.e. zero in L), and f evaluated a those have a multiplicative inverse $\frac{1}{f} + \mathfrak{m}$ in L, hence isn't zero in L.

It remains to transfer this L-solution to the system to a $\mathbb{C}$-solution to the system. There are 2 ways of doing this: the model theory ACF transfer principle (transferring the satisfaction of the 1st order ($\mathscr{L}_{\text{ring}}$-)sentence $\varphi \coloneqq \exists \mathsf{x}_1 \dots \mathsf{x}_d \big(\bigwedge_{i=1}^n g_i(\overline{\mathsf{x}}) \mathsf{=} 0 \wedge \neg f(\overline{\mathsf{x}}) \mathsf{=} 0 \big)$ between L and $\mathbb{C}$); or Zariski's [f.g. field algebraicity] lemma (which tells us in fact $L \cong \mathbb{C}$). Skip to whichever section you want. $\square$

4.2 Model theory ACF transfer principle

Again, in another section I will finish a self-contained proof of the Nullstellensatz without model theory, so you can skip to that if you don't want to read about model theory. This section will discuss essentially the main results of Chernikov Lec. 13. The main ideas are sketched in Lec. 12 and 13 of my 220A notes.

For terminology (like languages, structures, words, terms, formulas, free/bound variables, sentences, etc.), see the e.g. first 2 lectures of my 220A notes. They are somewhat long definitions, but largely trivial.

Recall first $\mathscr{L}_{\text{ring}} \coloneqq \{\underbrace{\mathsf{(}, \mathsf{)}, \exists, \forall, \neg, =, \wedge, \vee}_{\text{logical symbols}}, \text{formal variable symbols}\} \cup \{\mathsf{0}, \mathsf{1}, \mathsf{-}, \mathsf{+}, \mathsf{\cdot}\}$, and the axioms of a field can be written as one $\mathscr{L}_{\text{ring}}$-sentence φ_{field} formed by conjuncting the [identity / inverse / commutativity / associative] axioms of addition and multiplication, and the distributative law, and $\neg \mathsf{0=1}$. (For example, the mult. inverse axiom is $\forall \mathsf{x} (\neg (\mathsf{x=0}) \to \exists \mathsf{y} (\mathsf{x \cdot y = 1}))$.) We define property $P =$ "is a field" of $\mathscr{L}_{\text{ring}}$-structures to be the property axiomatized by φ_{field}.

Definition 4.1: ACF

The theory of algebraically closed fields is the $\mathscr{L}_{\text{ring}}$-theory, denoted by ACF, consisiting of φ_{field} and, for each $n \in \mathbb{N}^+$, a sentence ρ_n ("rho" for "root") expressing that any polynomial of degree n has a root, i.e. $\rho_n \coloneqq \forall \mathsf{z}_0 \dots \mathsf{z}_{n-1} \exists \mathsf{x} (\mathsf{x}^n \mathsf{+} \mathsf{z}_{n-1} \mathsf{x}^{n-1} \mathsf{+} \dots \mathsf{+} \mathsf{z}_0 \mathsf{=0})$ (we are using abbreviation $\mathsf{x}^n \coloneqq (\underbrace{\mathsf{x} \cdot \dots \cdot \mathsf{x}}_{n \text{ times}})$). (ACF is infinite axiomatization of the property "is an algebraically closed field", a property of $\mathscr{L}_{\text{ring}}$-structures)

Definition 4.2: Theory ACF_p

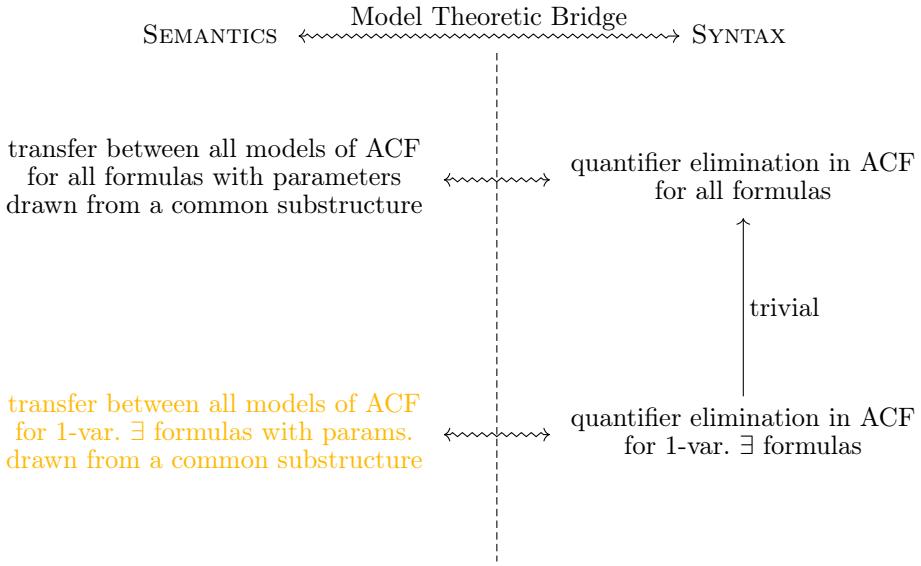
For p a prime number $\in \mathbb{N}_+$, let ACF_p be the theory ACF union the sentence $\chi_p := \{\underbrace{1+\ldots+1}_{p \text{ times}}=0\}$.
For $p = 0$, let ACF_0 denote $\mathrm{ACF} \cup \{\neg\chi_p : p \text{ prime number}\}$

Theorem 4.2: ACF_p complete

Let p be an prime ($\in \mathbb{N}$), or $p = 0$. Then ACF_p is complete.
In particular, every 1st order sentence φ in the language $\mathscr{L}_{\mathrm{ring}}$ is true in a ACF_p field $\iff$ φ is true in every ACF_p field.
In other words, think of this as a "transfer principle", transfering any 1st order truth from any ACF_p field to any other.

To give a taste of this result, I will prove the below GOAL statement. Amazingly, through the power of (some pretty substantial/"meaty") model theory, the GOAL statement is equivalent to the statement that ACF has quantifier elimination (QE), i.e. any $\mathscr{L}_{\mathrm{ring}}$-formula $\varphi(\mathsf{x}_1, \ldots, \mathsf{x}_d)$ is equivalent in ACF to some quantifier-free (q.f.) $\mathscr{L}_{\mathrm{ring}}$-formula $\psi(\mathsf{x}_1, \ldots, \mathsf{x}_d)$. From QE, one can very quickly prove Thm. 4.2 (by transfering truths through a common substructure — which one can always do with ACF; namely $\mathbb{F}_p$ is a common substructure of all ACF_p fields, and $\mathbb{Q}$ is a common substructure of all ACF_0 fields — which works because substructures/extensions preserve truth of quantifier free sentences).

The strategy is summarized by this graphic:



In the next 2 subsections, I will explain resp. the yellow GOAL statement (transfer 1-var. formulas), and then the Model Theoretic Bridge between syntax (QE) and semantics (model transference). Let me first quickly explain the "trivial" arrow in the graphic:

Lemma 4.1: Syntactic criterion (in fact characterization) for theory to have QE

Assume that for every q.f. formula φ and variable x (φ is truly arbitrary, can be in any number of variables, doesn't even have to mention x, though in that case $\exists \mathsf{x}\varphi$ trivially logically equivalent to the q.f. formula φ), there exists a q.f. formula ψ s.t. the formula $\exists \mathsf{x}\varphi$ is equivalent in T to ψ. Then T admits QE. Obviously converse also holds.

Proof: just induction on formula-height.

More explicitly, let ψ, ψ' be two formulas equivalent in T, denoted $\psi \sim_T \psi'$. Since (I proved this in Lemma in 220Ahw1p9) also $\neg\psi \sim_T \neg\psi', \exists \mathsf{x}\psi \sim_T \exists \mathsf{x}\psi'$ hence $\forall \mathsf{x}\psi \sim_T \forall \mathsf{x}\psi'$, and $\chi \wedge \psi \sim_T \chi \wedge \psi'$ (for any formula χ), we can argue by induction on $\mathrm{ht}(\psi)$: suffice to consider only formulas in prenex normal form $\varphi \coloneqq \mathsf{Q}_1\mathsf{x}_1 \dots \mathsf{Q}_n\mathsf{x}_n\chi$ (once we show those equiv. in T to q.f. formula, use transitivity of equivalence relation $\sim_T$), the formula $\mathsf{Q}_2\mathsf{x}_2 \dots \mathsf{Q}_n\mathsf{x}_n\chi$ is smaller height so by induction hypothesis $\sim_T$ to q.f. formula ψ, so then $\varphi \sim_T \mathsf{Q}_1\mathsf{x}_1\psi$. By assumption of this lemma, there is q.f. formula $\sim_T$ to $\mathsf{Q}_1\mathsf{x}_1\psi$ and hence to φ by transitivity. ■

Semantic transfer of one-variable existential formulas

- **Goal:** if K, L are alg. closed fields (i.e. 2 models of theory ACF), and A is common subring of K, L (i.e. a common $\mathscr{L}_{\mathrm{ring}}$-substructure), and $\varphi(\mathsf{x}_0, \dots, \mathsf{x}_d)$ is q.f. $\mathscr{L}_{\mathrm{ring}}$-formula, then for any $\overline{a} \in A^d$, [there existing $b \in L$ s.t. $L \models \varphi(b, \overline{a})$] $\implies$ there exists $c \in K$ s.t. $K \models \varphi(c, \overline{a})$.

  Note: because we quantify/range over all models K, L of ACF, the **Goal** is actually equivalent to putting "$\iff$" in place of "$\implies$" (by symmetry).

We can now make several reductions:

- **Claim1:** observe that if $A' \supseteq A$ is also common subring of K, L, it suffices to prove **Goal** for just A' since "$\forall \overline{a} \in A'$" $\rightsquigarrow$ "$\forall \overline{a} \in A$". The key idea is that by Fact (2) and (3), K and L both contain alg. closures of A, denoted F_K and F_L resp., that are isomorphic via some map ι that is identity on A, so up to identification of elements of K, L by ι, we can assume K, L contain a common alg. closed subfield $A' \coloneqq F_L \simeq F_K$ (without identifying elements by ι, this is the statement $\overline{a} \in F_L$, prove $\exists b \in L$ s.t. $L \models \varphi(b, \overline{a})$ implies $\exists c \in K$ s.t. $K \models \varphi(c, \iota(\overline{a}))$). That is to say, we have reduced to proving **Goal1**: the original **Goal** in the special case of *alg. closed* $A \subseteq K, L$.
- **Claim2:** in fact, it suffices to prove **Goal1** ("alg. closed $A_0 \subseteq K, L$") in the super-special case "$A \coloneqq K \subseteq L$" (this is now **Goal2**), since we can just apply the super-special case twice (once for "$A \coloneqq A_0 \subseteq K$"; and once for "$A \coloneqq A_0 \subseteq L$") to get: for all $\overline{a} \in A_0^d$, $\exists b \in L$ s.t. $L \models \varphi(b, \overline{a}) \iff \exists a_0 \in A_0$ s.t. $A_0 \models \varphi(a_0, \overline{a}) \iff \exists c \in K$ s.t. $K \models \varphi(c, \overline{a})$, where the two ends of this iff-chain being equivalent is exactly **Goal1** for the (abitrarily) chosen/given A_0.

We have simplified the semantic parts of the goal, but we can also simplify the syntactic part of the goal:

- **Claim (general observation about q.f. formulas in any language):** any q.f. formula like φ is logically equiv. to a formula in disjunction normal form, i.e. a formula of the form $\bigvee_i \bigwedge_j \chi_{i,j}$ where each $\chi_{i,j}(\mathsf{x}_0, \dots, \mathsf{x}_d)$ is either atomic or negated atomic (I also write this as $\bigvee_i \bigwedge_j \pm\chi_{i,j}$ for atomic $\chi_{i,j}$). In this form, φ is satsified iff at least one of its disjunctions is satisfied, so it suffices to prove **Goal2** in only the case that φ is a conjunction $\bigwedge$ of atomic or negated atomic formulas.

We are now done with all the reductions. In $\mathscr{L}_{\text{ring}}$, any atomic formula is equivalent to $p(\bar{\mathsf{x}})\mathsf{=0}$ for some polynomial p in variables $\bar{\mathsf{x}}$ and with *integer* coefficients. Thus, we have boiled everything down to showing:

- **FinishLine:** given an alg. closed field $K \subseteq L$ and an $\mathscr{L}_{\text{ring}}$-formula of the form $\bigwedge_{i=1}^{n} P_i(\bar{\mathsf{x}})\mathsf{=0} \wedge \bigwedge_{j=1}^{m} \neg Q_j(\bar{\mathsf{x}})\mathsf{=0}$ (with $P_i, Q_j \in \mathbb{Z}[\bar{\mathsf{x}}]$); THEN $\forall \bar{a} \in K^d$, if $[\exists b \in L$ s.t. on $(b, \bar{a})$, all P_i simultaneously vanish and all Q_j simultaneously do not vanish $\left\{\begin{smallmatrix}\text{all } P_i(b,\bar{a})=0\\ \text{all } Q_j(b,\bar{a})\neq 0\end{smallmatrix}\right.]$, then $[\exists c \in K$ s.t. on $(c, \bar{a})$, $\left\{\begin{smallmatrix}\text{all } P_i(c,\bar{a})=0\\ \text{all } Q_j(c,\bar{a})\neq 0\end{smallmatrix}\right.]$.

Indeed, if *some* $P_i(\mathsf{x}_0, a_1, \ldots, a_n) \in K[\mathsf{x}_0]$ is not the zero polynomial, everything is now super easy, because then for any $b \in L$ a root of the single variable polynomial $P_i(\mathsf{x}_0, \bar{a}) \in K[\mathsf{x}_0]$, b is algebraic over K, but K being algebraically closed tells us that actually b was already $\in K$ (see Fact (6)). I.e. as long as *some* $P_i(\mathsf{x}_0, \bar{a}) \in K[\mathsf{x}_0]$ is non-zero, then any $b \in L$ that satisfies the precondition $\left\{\begin{smallmatrix}\text{all } P_i(b,\bar{a})=0\\ \text{all } Q_j(b,\bar{a})\neq 0\end{smallmatrix}\right.$ automatically forces $b \in K$ and hence $c := b \in K$ satisfies the postcondition.

Thus the only remaining case is is $\varphi = \bigwedge_{j=1}^{m} \neg Q_i(\bar{\mathsf{x}})\mathsf{=0}$. By the assumed existence of $b \in L$ s.t. all $Q_j(b, \bar{a}) \neq 0$ (the LHS precondition), we get that each polynomial $Q_i(\mathsf{x}_0, \bar{a}) \in K[\mathsf{x}_0]$ is non-zero, hence has only a finite number of roots (single variable polys have $\leqslant$ deg many roots). But Fact (5) says K is infinite, meaning we can find $c \in K$ that is not a root of any $Q_j(\mathsf{x}_0, \bar{a})$ for $j \in 1..m$. And then indeed $K \models \varphi(c, \bar{a})$.

Syntax-semantics bridge (to boost transferability to all formulas)

We want to upgrade from transferring one-variable existential formulas to arbitrary variable existential sentences (which is enough to prove the Nullstellensatz). I present a model-theoretic way of doing this boost, by showing that the above transfer principle is equivalent to a syntactic condition called quantifier elimination. From the syntactic point of view, it then becomes trivial to boost from one-variable existential formulas to in fact arbitrary formulas.

The only tool we use in the proof (besides the all the terminology (see definitions in my notes or Hils-Loeser): language $\mathscr{L}$, theories T, consistent, proven $\vdash$ in T, equivalent in T, $\mathscr{L}$-structures, substructures, reduct, models of T, modelling $\models$ a formula, expansion by definition, etc.) is

Theorem 4.3: Gödel completeness theorem, model existence formulation

An $\mathscr{L}$-theory T has a model $\iff$ it is consistent.

I'll write it up, just so people get a sense of the ideas (even if you may not be familiar with the terminology; though again I promise it takes at most 1 hour to get a decent sense of all the terminology necessary, ignoring perhaps e.g. the finer details of the definition of the proof system and other such technical formalizations of things you really do understand well on an intuitive level). For a video version of what follows, see Chernikov Lec. 12.

Theorem 4.4: Equivalent definitions of quantifier elimination

Throughout this theorem/proof, just focus on $\mathscr{L} := \mathscr{L}_{\text{ring}}$-theory $T := \text{ACF}$.

Let T be any $\mathscr{L}$-theory, $n \in \mathbb{N}^+$, and fix any $\mathscr{L}_{\text{ring}}$-formula $\varphi(\mathsf{x}_1, \ldots, \mathsf{x}_n)$. Then, t.f.a.e.:

(1) φ is equivalent in T to some q.f. $\mathscr{L}$-formula $\psi(\mathsf{x}_1, \ldots, \mathsf{x}_n)$.

(2) for any two models $\mathcal{M}, \mathcal{N}$ of T, any common $\mathscr{L}$-substructure $\mathcal{A} \subseteq \mathcal{M}, \mathcal{N}$, and all $\overline{a} \in A^n$, we have that $\mathcal{M} \models \varphi(\overline{a}) \iff \mathcal{N} \models \varphi(\overline{a})$. It is perhaps more natural to instead say "common substructure up to isomorphism", since the specific base sets don't matter, merely the relationship between the elements as encoded by the $\mathscr{L}$-structures in question. Ultimately it doesn't matter because although my reformulation looks stronger at face-value, if we have $\mathcal{A}, \mathcal{A}'$ substructures of $\mathcal{M}, \mathcal{M}'$ resp. and $\mathcal{A} \cong \mathcal{A}'$, then we can find $\widetilde{\mathcal{M}}, \widetilde{\mathcal{M}}'$ isomorphic to $\mathcal{M}, \mathcal{M}'$ resp. s.t. $\widetilde{\mathcal{M}}, \widetilde{\mathcal{M}}'$ have base sets that actually literally contain the base set A, thus reducing to (2) as written.

In other words, this theorem tells us the only way truth of φ could be "independent" of the model of T — i.e. $\varphi(a_1, \dots, a_n)$ true in all models of T in which a copy of the elements $a_1, \dots, a_n$, related to each other in some fixed predetermined way (as encoded by the $\mathscr{L}$-structure $\mathcal{A}$), exist — is if φ is (logically equivalent in T to) a q.f. formula.

Remark: what if $n = 0$, i.e. φ is a sentence? Can consider φ as a formula $\varphi(\mathtt{x})$ and apply Thm to get a q.f. formula $\psi(\mathtt{x})$, which is equivalent in T to $\varphi(\mathtt{x})$. Example: if $\varphi := \mathtt{\exists y (y=y)}$ (which is provably from the empy theory, see Lec. 5), then we can take $\psi(\mathtt{x}) := \mathtt{x=x}$ because indeed $\varnothing \vdash \mathtt{\forall x (\exists y (y=y) \mapsto x=x)}$.

Note also if $\mathscr{L}$ has no constant symbols, then there are no quantifier free $\mathscr{L}$-sentences (just by syntactic rules of sentence formation). In this case, (in the course of the proof of Thm) when we assert the existence of q.f. formula ψ equiv to sentence φ (in the sense of the Rmk. above), we allow ψ to have a single free variable.

Another way to address this isue, is to require *all* languages to possess some logical (constant) symbol, e.g. `True` and `False`.

My warning: I mistakenly thought (2) above is equivalent to "for all models $\mathcal{A} \models T$ and all $\overline{a} \in A^n$, $\varphi(\overline{a})$ is sasifed in $\mathcal{A}$ iff satisfied in any extension of $\mathcal{A}$". This is not right, because (2) does not need $\mathcal{A}$ to be model of T; it says substructure, not elementary substructure!

Proof (of Thm. 4.4): (1) $\implies$ (2): any formula $\varphi(\overline{\mathtt{x}})$ is equivalent in T to a q.f. formula $\psi(\overline{\mathtt{x}})$, so for any $\mathscr{L}$-substructure and $\mathscr{L}$-models $\mathcal{A} \subseteq \mathcal{M}, \mathcal{N} \models T$ and any $\overline{a} \in A^\bullet$, we have $\mathcal{M}, \mathcal{N} \models \mathtt{\forall \overline{x} (}\varphi \mapsto \psi\mathtt{)} \implies$ for any $\overline{a} \in A^\bullet$,

$$\mathcal{M} \models \varphi(\overline{a}) \iff \mathcal{M} \models \psi(\overline{a}) \iff \mathcal{A} \models \psi(\overline{a})$$

where we used that ψ q.f. means any extension/substructure preserves truth value (220Ahw1p7.1). The same is true for $\mathcal{N}$ in place of $\mathcal{M}$, so indeed

$$\mathcal{M} \models \varphi(\overline{a}) \iff \mathcal{M} \models \psi(\overline{a}) \iff \mathcal{A} \models \psi(\overline{a}) \iff \mathcal{N} \models \psi(\overline{a}) \iff \mathcal{N} \models \varphi(\overline{a}).$$

(2) $\implies$ (1): our goal is to find a q.f. formula $\chi^\star$ s.t. "$T \vdash (\chi^\star \mapsto \varphi)$". We define the "haystack" Γ, in which we search for our desired needle $\chi^\star$: $\Gamma(\overline{\mathtt{x}}) := \{\chi(\mathtt{x}_1, \dots, \mathtt{x}_n)$ q.f. $\mathscr{L}$-fmls. $: T \models \mathtt{\forall x}_1 \dots \mathtt{x}_n (\varphi \to \chi)\}$ ("the set of all q.f. $\mathscr{L}$-formulas $\chi(\overline{\mathtt{x}})$ with $\leqslant n$ free variables that are implied in T by φ", i.e. the set of all q.f. $\chi(\overline{\mathtt{x}})$ s.t. $\forall \mathcal{M} \models T$ and $\forall$ 'satisfiers' $\overline{s} \in M^n$ s.t. $\mathcal{M} \models \varphi(\overline{s})$, also satisfy $\mathcal{M} \models \chi(\overline{s})$), so that our goal can be rephrased as finding some $\chi^\star(\overline{\mathtt{x}}) \in \Gamma(\overline{\mathtt{x}})$ to conversely imply φ in T, or equivalently (by compactness) $T \cup \Gamma(\overline{\mathtt{x}}) \models \varphi(\overline{\mathtt{x}})$.

We introduce new constants in place of $\overline{\mathtt{x}}$ ("simulating $\overline{\mathtt{x}}$") because it's easier to work with those to create models — in particular we defined theories to consist of sentences, not formulas: let $\mathtt{c}_1, \dots, \mathtt{c}_n$ be new pairwise distinct constant symbols, and define $\Gamma(\overline{\mathtt{c}}) := \{\chi(\mathtt{c}_1, \dots, \mathtt{c}_n) : \chi \in \Gamma(\overline{\mathtt{x}})\}$ (i.e. the substitution of $\mathtt{c}_i$ for $\mathtt{x}_i$ in the formulas χ), which is then a theory in the language $\mathscr{L}' := \mathscr{L} \cup \{\mathtt{c}_1, \dots, \mathtt{c}_n\}$.

- **Claim:** the $\mathscr{L}'$-theory $T \cup \Gamma(\overline{\mathsf{c}})$ proves/models $\models \varphi(\overline{\mathsf{c}})$ (a $\mathscr{L}'$-sentence). ($T \cup \Gamma(\overline{\mathsf{c}})$ has a $\mathscr{L}'$-model, because by assumption (2) in Thm, we can assume we have an $\mathscr{L}$-model $\mathfrak{N} \models T$ satisfying $\mathfrak{N} \models \varphi(\overline{s})$ (for 'satisfiers' $\overline{s} \in \mathfrak{N}^n$), which implies $\mathfrak{N} \models \chi(\overline{s})$ for all $\chi(\overline{\mathsf{x}}) \in \Gamma(\overline{\mathsf{x}})$, by def. of Γ. Thus, interpreting $\mathsf{c}_i^{\mathfrak{N}'} := s_i \in \mathfrak{N}$, we get $\mathfrak{N}' \models T \cup \Gamma(\overline{\mathsf{c}})$.)
- *Proof (spread out in following ▶):* suppose not, $\not\models \varphi(\overline{\mathsf{c}})$. Then, by the completeness theorem,
- $\exists$ an $\mathscr{L}'$-model $\mathcal{M}' \models T \cup \Gamma(\overline{\mathsf{c}}) \cup \{\neg\varphi(\overline{\mathsf{c}})\}$.
- Let $\mathcal{A}' := \langle \mathsf{c}_1^{\mathcal{M}'}, \ldots, \mathsf{c}_n^{\mathcal{M}'} \rangle_{\mathcal{M}'}$ be the $\mathscr{L}'$-substructure of $\mathcal{M}'$ generated by the interpretations of the new constant symbols $\mathsf{c}_1^{\mathcal{M}'}, \ldots, \mathsf{c}_n^{\mathcal{M}'} \in \mathcal{M}'$ — the $\langle \bullet \rangle_{\mathcal{M}'}$ notation defined at end of Lec9 ('Class12').
- Let A denote the base set of $\mathcal{A}'$. In particular, note that
- all $a_i := \mathsf{c}_i^{\mathcal{M}'} \in A$.
- The STRATEGY, is to build an $\mathscr{L}'$-model $\mathcal{N}^* \models T \cup \{\varphi(\overline{\mathsf{c}})\}$ that has an $\mathscr{L}'$-substructure isomorphic to $\mathcal{A}'$. This reaches a contradiction, because we will get two $\mathscr{L}$-structures $\mathcal{N} := \mathcal{N}^*\restriction_{\mathscr{L}}$, $\mathcal{M} := \mathcal{M}'\restriction_{\mathscr{L}}$ both modelling T, and both containing (up to iso.) the common $\mathscr{L}$-substructure $\mathcal{A} := \mathcal{A}'\restriction_{\mathscr{L}}$, in particular both containing $a_i := \mathsf{c}_i^{\mathcal{M}'} \in A$, which satisfy $\mathcal{N} \models \varphi(\overline{a})$ but $\mathcal{M} \models \neg\varphi(\overline{a})$; contradiction to assumption (2) in Thm.
  We can ensure the $\mathscr{L}'$-model $\mathcal{N}^*$ contains (an isomorphic copy of) $\mathcal{A}'$ by asking $\mathcal{N}^*$ to also model the simple diagram (of the $\mathscr{L}$-reduct) $\Delta(\mathcal{A}'\restriction_{\mathscr{L}})$, because by definition of simple diagram Δ, any $\mathscr{L}'$-structure $\mathcal{N}^*$ that models $\Delta(\mathcal{A}'\restriction_{\mathscr{L}})$ contains an isomorphic copy of $\mathcal{A}'$ as an $\mathscr{L}'$-substructure.
- Let us refresh ourselves on simple diagrams, by making the exercise observation that our haystack ("substituting $\overline{\mathsf{c}}/\overline{\mathsf{x}}$ in the set of all q.f. $\mathscr{L}$-formulas $\chi(\overline{\mathsf{x}})$ that are implied in T by φ") $\Gamma(\overline{\mathsf{c}}) \subseteq \Delta(\mathcal{A}'\restriction_{\mathscr{L}})$, the simple diagram of the $\mathscr{L}$-reduct $\mathcal{A}'\restriction_{\mathscr{L}}$ — which recall (from end of Lec. 10) is the set of all $\mathscr{L}_A$-sentences $\phi(\mathsf{c}_{m_1}, \ldots, \mathsf{c}_{m_?})$ (with ϕ a q.f. $\mathscr{L}$-fml.) satisfied in $\mathscr{L}_A$-str.: $\langle A; \ldots \rangle$ with interpretations of $\mathscr{L}$-symbols coming from $\mathcal{A}'\restriction_{\mathscr{L}}$, and new constant symbols $\mathsf{c}_m \in \mathscr{L}_A \setminus \mathscr{L}$ interpreted to $m \in A$. Our current c_i would be written in the old notation as $\mathsf{c}_{(\mathsf{c}_i^{\mathcal{M}'})}$, giving an inclusion $\mathscr{L}' \hookrightarrow \mathscr{L}_A$. The containment ("$\Gamma \subseteq \Delta$") is true because

$$\mathcal{M}' \models \Gamma(\overline{\mathsf{c}}) \text{ (q.f. } \mathscr{L}'\text{-fmls.!) and } \mathcal{A}' \subseteq \mathcal{M}' \implies \mathcal{A}' \models \Gamma(\overline{\mathsf{c}}) \iff \mathcal{A}'\restriction_{\mathscr{L}} \models \Gamma(\mathsf{c}_1^{\mathcal{M}'}, \ldots, \mathsf{c}_n^{\mathcal{M}'}).$$

  Henceforth, the notation $\Delta(\mathcal{A}')$ means $\Delta(\mathcal{A}'\restriction_{\mathscr{L}})$ (typecheckers recall: it's a set of q.f. $\mathscr{L}_A$-sentences).
- As outlined above, the proof hinges upon the following **Subclaim** ("constructing $\mathcal{N}^*$ opposing $\mathcal{M}'$"):
  - **Subclaim:** the theory $\Sigma := T \cup \Delta(\mathcal{A}') \cup \{\varphi(\overline{\mathsf{c}})\}$ has an $\mathscr{L}_A$-model.
    * Suppose not. Then every $\mathscr{L}_A$-model of $T \cup \Delta(\mathcal{A}')$ satisfies $\neg\varphi(\overline{\mathsf{c}})$, so by Gödel completeness (or Cor. ?? of Deduction Lemma in Lec. 5) $T \cup \Delta(\mathcal{A}') \vdash \neg\varphi(\overline{\mathsf{c}})$.
    * Now, every element $a \in A$ can be written as an $\mathscr{L}'$-term (without free variables) $t_a^{\mathcal{M}'}(\overline{\mathsf{c}})$ (from the explicit description of base sets of generated structures from Lec. 9: the generated substructure $\mathcal{A}' := \langle a_1, \ldots, a_n \rangle_{\mathcal{M}'} = \bigcap_{\{a_1,\ldots,a_n\} \subseteq \mathfrak{N} \subseteq \mathcal{M}'} \mathfrak{N}$ has base set

$$\{\text{interpretations } t^{\mathcal{M}'}(\overline{a}) : \mathscr{L}'\text{-terms } t(\overline{\mathsf{x}}) \in \mathscr{T}^{\mathscr{L}'}, \overline{a} \in \{a_1, \ldots, a_n\}^?\}.$$

      For example for $\mathscr{L} = \mathscr{L}_{\text{ring}}$ and $T = \text{ACF}$, substructure = subring, and subring generated by $a_1, \ldots, a_n$ in ACF $\mathcal{M}'$ has base set given by all finite combinations of $a_1, \ldots, a_n$ by $+, -, \cdot$ = interpretations of $\mathscr{L}_{\text{ring}}$ terms).
    * Denote by $\Delta_{\overline{\mathsf{c}}}(\mathcal{A}')$ the set of q.f. $\mathscr{L}'$-sentences in $\Delta(\mathcal{A}')$. Note $\Gamma(\overline{\mathsf{c}}) \subseteq \Delta_{\overline{\mathsf{c}}}(\mathcal{A}') \subseteq \Delta(\mathcal{A}')$.
    * We have that the $\mathscr{L}_A$-theory $T \cup \Delta(\mathcal{A}')$ is a conservative expansion of $\mathscr{L}'$-theory $T \cup \Delta_{\overline{\mathsf{c}}}(\mathcal{A}')$

(recall Lec. 11, means that both theories exactly agree on provability of an $\mathscr{L}'$-sentence); one way to see this (which works for general T) is from the syntactic viewpoint: the former is essentially an expansion by definition of the latter (meaning the only new sentences in larger theory are those literally *defining* the new relation/function/constant symbols in the larger language in terms of formulas in the smaller language). Another way (in our specific $T = \mathrm{ACF}$ setting) is the semantic viewpoint:

- Fix any ACF M', and $a_1, \ldots, a_n \in M'$. Define $A' \subseteq M'$ to be the subring generated by $a_1, \ldots, a_n \in M'$. (To match how we started this proof with $\mathcal{M}'$ and $\mathcal{A}'$.)
- Let $\mathfrak{N}$ be another ACF, with $b_1, \ldots, b_n \in \mathfrak{N}$ satisfying exactly the same $\mathscr{L}_{\mathrm{ring}}$ relationships as $a_1, \ldots, a_n \in M'$. In particular, $\Psi(a_i) \coloneqq b_i$ should define a ring homomorphism (in fact isomorphism! since the property of being non-zero is preserved $\rightsquigarrow$ injectivity) from A' to the generated subring $\langle b_1, \ldots, b_n \rangle \subseteq \mathfrak{N}$. I.e. $\mathfrak{N}$ contains an isomorphic copy of A'.
- We have thus shown that any $\mathscr{L}'$-model $\mathfrak{N} \models \mathrm{ACF} \cup \Delta_{\overline{\mathsf{c}}}(\mathcal{A}')$ is also an $\mathscr{L}_A$-model $\models \mathrm{ACF} \cup \Delta(\mathcal{A}')$. This proves the desired conservative expansion property.

$*$ Then in particular (using def. of conservative exp.), $T \cup \Delta_{\overline{\mathsf{c}}}(\mathcal{A}') \vdash \neg\varphi(\overline{\mathsf{c}})$.

$*$ To summarize, although may have terms involving constants not in $\overline{\mathsf{c}} = (\mathsf{c}_1, \ldots, \mathsf{c}_n) = (\mathsf{c}_{a_1}, \ldots, \mathsf{c}_{a_n})$ in $\Delta(\mathcal{A}')$ in the proof $T \cup \Delta(\mathcal{A}') \vdash \neg\varphi(\overline{\mathsf{c}})$, we use that all such terms can be written in terms of $\overline{\mathsf{c}}$, so the smaller theory already proves it: $T \cup \Delta_{\overline{\mathsf{c}}}(\mathcal{A}') \vdash \neg\varphi(\overline{c})$.

$*$ By compactness, there exists q.f. $\mathscr{L}$-formulas $\xi_1(\overline{\mathsf{x}}), \ldots, \xi_k(\overline{\mathsf{x}})$ s.t. all $\xi_i(\overline{\mathsf{c}}) \in \Delta_{\overline{\mathsf{c}}}(\mathcal{A}')$ and $T \cup \{\xi_i(\overline{\mathsf{c}})\}_{i=1}^k \vdash \neg\varphi(\overline{\mathsf{c}})$, which is iff (by Deduction Lemma of Lec. 5, and trivial conjunction step)

$$T \models \underbrace{\textstyle\bigwedge_{i=1}^k \xi_i(\overline{\mathsf{c}})}_{=:\xi(\overline{\mathsf{c}})} \to \varphi(\overline{\mathsf{c}}).$$

I point out that obviously, $\Delta_{\overline{\mathsf{c}}}(\mathcal{A}') \models \xi(\overline{\mathsf{c}})$. (We will later contradict this.)

$*$ However, note that the constant symbols $\mathsf{c}_1, \ldots, \mathsf{c}_n$ don't appear in T, nor in $\varphi(\overline{\mathsf{x}})$ nor $\xi(\overline{\mathsf{x}})$. Thus, using Simulation of Contants by Variables lemma (Lec. 6), we get that $T \models (\xi(\overline{\mathsf{x}}) \to \neg\varphi(\overline{\mathsf{x}}))$, and by (Gen), $T \models \forall\overline{\mathsf{x}}(\xi(\overline{\mathsf{x}}) \to \neg\varphi(\overline{\mathsf{x}}))$.

$*$ Taking contrapositive, $T \models \forall\overline{\mathsf{x}}(\varphi(\overline{\mathsf{x}}) \to \neg\xi(\overline{\mathsf{x}}))$.

$*$ As all $\xi_i(\overline{\mathsf{x}})$ were q.f., $\neg\xi_i(\overline{\mathsf{x}})$ is too, so $\neg\xi(\overline{\mathsf{x}}) \in \Gamma(\overline{\mathsf{x}})$, so $\neg\xi(\overline{\mathsf{c}}) \in \Gamma(\overline{\mathsf{c}}) \subseteq \Delta(\mathcal{A}')$.

$*$ This contradicts what we said earlier: $\Delta_{\overline{\mathsf{c}}}(\mathcal{A}') \vdash \xi(\overline{\mathsf{c}}) \implies \Delta(\mathcal{A}') \vdash \xi(\overline{\mathsf{c}})$. This is a genuine contradiction because we know $\Delta(\mathcal{A}')$ is consistent (it's a subset of the complete diagram, which is Th(a model), hence always complete $\implies$ consistent).

– So, we have shown the **Subclaim**, i.e. $\Sigma \coloneqq T \cup \Delta(\mathcal{A}') \cup \{\varphi(\overline{c})\}$ has a model.

Tidying up/summarizing and fulfilling our promises:

▶ Returning to the original **Claim:** now we know the $\mathscr{L}'$-theory Σ has a $\mathscr{L}'$-model $\mathcal{N}^*$, and the $\mathscr{L}$-reduct $\mathcal{N} \coloneqq \mathcal{N}^*\restriction_{\mathscr{L}}$ contains an isomorphic copy $\mathcal{B}'$ of $\mathcal{A}'$, because in particular $\mathcal{N}^*$ is model of $\Delta(\mathcal{A}') \subseteq \Sigma$ (and any model of simple diagram of a structure contains an isomorphic copy of that structure as substructure; also going to $\mathscr{L}$-reduct only throws away irrelevant interpretations in this case).

So up to identification of $\mathcal{B}'$ and $\mathcal{A}'$, we have constructed two models $\mathcal{M} \coloneqq \mathcal{M}'\restriction_{\mathscr{L}}$ and $\mathcal{N}$ of T, containing a common substructure $\mathcal{A} \coloneqq \mathcal{A}'\restriction_{\mathscr{L}}$, s.t. if we set $a_i \coloneqq \mathsf{c}_i^{\mathcal{M}'}$, then defining $\overline{a} \coloneqq (a_i)_1^n$ we get $\mathcal{N} \models \varphi(\overline{a})$ ("a sentence in theory Σ"), but recall from first line of **Claim** that by definition

$\mathcal{M}' \models \neg\varphi(\overline{a})$, which contradicts property (2) in the theorem. So, the **Claim:** $T \cup \Gamma(\overline{c}) \models \varphi(\overline{c})$ is proven.

Finally, by compactness, there exists $\xi_1(\overline{\mathsf{c}}), \ldots, \xi_m(\overline{\mathsf{c}}) \in \Gamma(\overline{\mathsf{c}})$ s.t. (similar to before)

$$T \models \bigwedge_{i=1}^{n} \xi_i(\overline{\mathsf{c}}) \to \varphi(\overline{\mathsf{c}}), \text{ and again as before, } T \models \forall \overline{\mathsf{x}} \big(\underbrace{\textstyle\bigwedge_{i=1}^{n} \xi_i(\overline{\mathsf{x}})}_{=:\xi(\overline{\mathsf{x}})} \to \varphi(\overline{\mathsf{x}}) \big).$$

From definition of $\Gamma(\overline{\mathsf{c}})$, we know $T \models \forall \overline{\mathsf{x}}(\varphi \to \xi_i)$ for $i \in [m]^+$, so we have $T \models \forall \overline{\mathsf{x}}(\xi(\overline{\mathsf{x}}) \leftrightarrow \varphi(\overline{\mathsf{x}}))$ where all ξ_i are q.f. $\mathscr{L}$-formulas $\rightsquigarrow \xi$ is q.f. $\mathscr{L}$-formula. This ends the proof of Thm. 4.4. ■

4.3 Zariski's lemma (end of self-contained full proof of Nullstellensatz)

Lemma 4.2: Zariski's (finitely generated) field algebraicity lemma

Let L be a field extension of a field k that is f.g. as a k-algebra (i.e. $L = k[\alpha_1, \ldots, \alpha_n]$ for some $\alpha_1, \ldots, \alpha_n \in L$). Then L is algebraic over k (i.e. all α_i are algebraic over k).

Preliminary/philosophical remarks

Allcock exposits his proof quite well. No fancy technology/concepts, very concrete (a bit computational...); things are first explained intuitively and then rigorously. There is literally no way to improve upon it. Lay your eyes upon its full majesty `https://web.ma.utexas.edu/users/allcock/expos/nullstellensatz3.pdf`...

MO (+ Keith Conrad comments and reference to Fulton's book *Algebraic Curves* (Chp.1 §10 Prop. 4)) informs me Azarang has also written up a short proof, which I think is **exactly the same** as Allcock's proof, but phrased slightly differently like using induction and using fancier language (integral extension). [Azarang's proof for $n > 1$ (see below) AFTER applying induction hypothesis, is the same as Allcock trancendence degree 1 case].

It is interesting to compare Allcock and Azarang proofs, and also the more well-known proofs using Noether normalization. Zanachan presents a/the proof via Noether normalization in 5 minutes.

In particular, Noether normalization gives a **direct proof**: we (assuming algebraic dependence, and a variable change trick — D&F pg.699) actually find a monic polynomial with coefficients in $k[\alpha_1, \ldots, \alpha_{n-1}]$ that α_n is a root of; finish the proof by induction and transitivity of integrality (overall, I think can unwind this to get an explicit polynomial out — (transitivity of) integrality intimately tied to determinant trick; kind of gross but still constructive).

Whereas Allcock/Azarang both prove by the **contrapositive** (Allcock)/**contradiction** (Azarang): we assume α_1 is NOT algebraic, so all polynomials in α_1 are not zero and can be inverted in the fraction field.

I was actually able to twist Allcock/Azarang's proof to make it constructive, by keeping careful track of the fractions involved.

Azarang's proof of Zariski's [f.g. field algebraicity] lemma

Zanachan and Azarang both use/isolate/distill the following baby lemma (in fact iff, see Thm. 2.7 in KConrad article):

Lemma 4.3: Fields can only integrally extend other fields

Suppose a field F is integral over a subdomain D. Then D must be a field.

Proof (taken from Zanachan slide): given nonzero $0 \neq f \in D$, want to show $f^{-1} \in D$ (where f^{-1} refers to the inverse of f in F). We assumed F is integral over D, so we have a monic polynomial relation ($n \geqslant 1, d_i \in D$):

$$(f^{-1})^m + d_{m-1}(f^{-1})^{m-1} + \ldots + d_0 = 0.$$

Multiplying by f^{m-1} and rearranging yields of course $f^{-1} \in D$. ■

We are now ready to prove Zariski's [f.g. field algebraicity] lemma, by contradiction. ...because [Azarang's proof for $n > 1$ (see below) AFTER applying induction hypothesis, is the same as Allcock trancendence degree 1 case], can actually stop using induction, and then steal Allcock's trancendence degree > 1 case to finish — allowing us to prove the contrapositive, instead of proof by contradiction. However, that's still not constructive/intuitionistic.

Proof: by induction on the number of generators $\alpha_1, \ldots, \alpha_n$ ($\in$ the larger field L) it takes to get from the base field k to to the larger field L.

For $n = 1$, we assume $L = k[\alpha]$.

Then L is a field implies $\alpha^{-1} \in L$, i.e. $\alpha^{-1} = \text{poly}(\alpha)$. Multiplying by α and rearranging, we get that α is a root of a polynomial $\in k[t]$.

For $n > 1$, we have $L = k[\alpha, \beta_2, \ldots, \beta_n]$, which also equals $= k(\alpha)[\beta_2, \ldots, \beta_n]$ ($\subseteq$ is obvious; and $\supseteq$ because L is a field).

Applying the induction hypothesis to larger field L over base $k(\alpha)$, we know $\beta_2, \ldots, \beta_n$ are algebraic over $k(\alpha)$, i.e. we have polynomials $f_j \in k(\alpha)[t]$ having β_j as a root (so to be explicit, coefficients of f_j are fractions with numerator and denominator $\in k[\alpha]$). More visually:

$$f_j = \frac{\text{poly} \in k[\alpha]}{\text{poly} \in k[\alpha]} \cdot t^{\star} + \frac{\text{poly} \in k[\alpha]}{\text{poly} \in k[\alpha]} \cdot t^{\star - 1} + \ldots + \frac{\text{poly} \in k[\alpha]}{\text{poly} \in k[\alpha]} \cdot t^0 \in k(\alpha)[t].$$

Let $c_j(\alpha) \in k[\alpha]$ be product of all numerator and denominators of coefficients of f_j (suffices to take product (in $k[\alpha]$) of numerator of leading order (in t) term of f_j, and denominators of all other coefficients), so that β_j is integral over $k[\alpha][\frac{1}{c_j(\alpha)}]$. We are allowed to divide by $c_j(\alpha)$ because we assume f.s.o.c. α is not algebraic over k.

Combining everything, get all β_j integral over $D \coloneqq k[\alpha][\frac{1}{c_2(\alpha)}, \ldots, \frac{1}{c_n(\alpha)}] \subseteq k(\alpha)$, so L integral over D. Baby Lemma 4.3 tells us D must be a field, i.e. $D \subseteq k(\alpha)$ actually has to $= k(\alpha)$.

But $k(x)$ is not a finitely generated $k[x]$-algebra, just like $\mathbb{Q}$ is not a f.g. $\mathbb{Z}$-algebra — both follow from infinitely many prime elements (e.g. Euclid's infinitude of primes argument). ■

Unfortunately I just realized Azarang's proof uses that having integral generating elements means you have an integral extension, which typically requires the Determinant Trick. So slightly technical. How does Allcock get around this?

Making Allcock's proof of Zariski's [f.g. field algebraicity] lemma constructive

Proof: same beginning as Azarang, up to "For $n > 1$, we have $L = k[\alpha, \beta_2, \ldots, \beta_n]$, which also equals $= k(\alpha)[\beta_2, \ldots, \beta_n]$... where $\beta_2, \ldots, \beta_n$ are algebraic over $k(\alpha)$"

From here, we consider $e_1, \ldots, e_\ell$ a (finite!) basis of L as a $k(\alpha)$-VS (i.e. e_i is a (culled if necessary) enumeration of products of $\beta_\bullet$; finitely many span L because $\beta_\bullet$ are algebraic over $k(\alpha)$). We write down the multiplication table for these basis elements:

$$e_i e_j = \sum_k \frac{p_{ijk}(\alpha)}{q_{ijk}(\alpha)} \cdot e_k,$$

where $p_{ijk}, q_{ijk} \in k[t]$ and $q_{ijk}(\alpha) \neq 0$ (the $k(\alpha)$ coefficients in terms of the basis $\{e_\bullet\}$ are well-defined).

Consider arbitrary $x \in L = k[\alpha, \beta_2, \ldots, \beta_n]$; i.e. x is a (finite!) k-linear combination (LC) of products of $\alpha, \beta_2, \ldots, \beta_n$, which are all themselves $k(\alpha)$-LCs of the basis $\{e_\bullet\}$:

$$\beta_j = \sum_k \frac{r_{jk}(\alpha)}{s_{jk}(\alpha)} \cdot e_k,$$

where $r_{jk}, s_{jk} \in k[t]$ and $s_{jk}(\alpha) \neq 0$. (Yes, most of these r, s polynomials are $\equiv 1$ or $\equiv 0$; unless β_j was culled when we extracted the basis $\{e_\bullet\}$ from the list of products of the $\beta_\bullet$.)

Plugging these $\{e_\bullet\}$ representations of β_j into the k-LC representing x, and using the multiplication table, we get

$$x = \sum_k \frac{P_x(\alpha)}{\prod q_\bullet(\alpha) \prod s_\bullet(\alpha)} \cdot e_k,$$

for $P_x \in k[t]$ (may depend on x). Call this big product denominator polynomial $=: Q \in k[t]$ — $Q(\alpha) \neq 0$.

We have shown that every $k(\alpha)$-LC of the basis $\{e_\bullet\}$ equals a $k[\alpha][\frac{1}{Q(\alpha)}]$ -LC, so by uniqueness of coefficients of LCs w.r.t. a basis, we conclude

$$k(\alpha) = k[\alpha]\left[\frac{1}{Q(\alpha)}\right].$$

To finish, let $Z \in k[t]$ relatively prime to $Q \in k[t]$ (can use Euclid infinite of primes trick). Then we know (if $Z(\alpha) = 0$, immediately done; else:)

$$\frac{1}{Z(\alpha)} = \frac{P(\alpha)}{Q(\alpha)} \implies Q(\alpha) - P(\alpha)Z(\alpha) = 0.$$

The LHS polynomial $Q(t) - P(t)Z(t)$ can not be identically $\equiv 0$, since that would imply $Q(t) = P(t)Z(t)$ in $k[t]$, but we demanded $Z(t)$ to be relatively prime to $Q(t)$. ■

This combines the induction of Azarang's proof, with Allcock's basis techniques and avoidance of fancy technology/terminology (like integral extensions), and my own twist to make it constructive:

- given $k(\alpha)[\beta_2, \ldots, \beta_n]$, where I know the $k(\alpha)$-polynomial with β_j as zero (actually just the degree suffices),
- I can form a finite spanning set of products of $\beta_\bullet$,
- which I can cull to get a basis $\{e_\bullet\}$,

- w.r.t. which I can write out LC representations for β_j, and also a multiplication table,
- whose coefficients' denominator polynomials I can multiply together to get $Q \in k[t]$,
- to which I can find $Z \in k[t]$ relatively prime and $P \in k[t]$ such that $\frac{1}{Z(\alpha)} = \frac{P(\alpha)}{Q(\alpha)}$,
- thus giving me the final explicit non-zero polynomial $Q - ZP \in k[t]$ which has α as a root.

I am finally satisfied!