CDM

Semirings

Klaus Sutner
Carnegie Mellon University

30-semi-rings 2017/12/15 23:16

A semiring is a structure $\langle X, \oplus, \otimes, 0, 1 \rangle$ that satisfies the following conditions:

1. $\langle X, \oplus, 0 \rangle$ and $\langle X, \otimes, 1 \rangle$ are monoids, the former is commutative.
2. Operation $\otimes$ distributes over $\oplus$ on the left and right: $x \otimes (y \oplus z) = (x \otimes y) \oplus (x \otimes z)$ and $(y \oplus z) \otimes x = (y \otimes x) \oplus (z \otimes x)$.
3. 0 is a null (or annihilator) with respect to $\otimes$: $x \otimes 0 = 0 \otimes x = 0$.

A semiring is commutative if $x \otimes y = y \otimes x$ and idempotent if $x \oplus x = x$.

Example

Some standard examples of semirings are the following. The ring of integers $\langle \mathbb{Z}, +, *, 0, 1 \rangle$ is a semiring under the standard operations of addition and multiplication. More generally any ring is a semiring; the additive monoid here is a group.

Example

The Boolean semiring $\mathbb{B} = \langle \{0, 1\}, \vee, \wedge, 0, 1 \rangle$ where the operations are logical 'or' and 'and'.

Example (The Relation Semiring)

The semiring $\mathcal{R}$ of all binary relations over a set A has set union as the additive operation, and relational composition as the multiplicative operation. 0 is the empty relation, and 1 is the identity relation.

Example (The Language Semiring)

The semiring $\mathcal{L}(\Sigma)$ of all languages over Σ is the algebra $\langle \mathbb{P}(\Sigma^\star), \cup, \cdot, \emptyset, \varepsilon \rangle$. It is easy to check that all the equations for a semiring are satisfied by $\mathcal{L}(\Sigma)$. Note that one can introduce a metric on $\mathcal{L}(\Sigma)$ by setting $dist(L, K) := 2^{-n}$ where n is minimal such that $L \cap \Sigma^n \neq K \cap \Sigma^n$ for $L \neq K$ and $dist(L, L) = 0$. It is not hard to see that $\mathcal{L}(\Sigma)$ is a complete metric space with respect to this distance function.

Example (Tropical Semiring)

The tropical semiring is defined by $TS = \langle \mathbb{N} \cup \{\infty\}, \min, +, \infty, 0 \rangle$. It is used for example in Dijkstra's algorithm for minimal cost paths. We will encounter it in chapter **??** in our discussion of the finite power property.

Example (Matrix Semirings)

Let $S = \langle S, \oplus, \otimes, 0, 1 \rangle$ be a semiring. The operations $\oplus$ and $\otimes$ naturally extend to n by n matrices over S. One can show that $\langle S^{n,n}, \oplus, \otimes, \mathbf{0}, \mathbf{1} \rangle$ is again a semiring. Here $\mathbf{0}$ and $\mathbf{1}$ are the appropriate null and identity matrices over S.

Matrix semirings will be of importance to us later. To describe the behavior of a finite state machine one can often use homomorphisms of the form $h : \Sigma^\star \to S^{n,n}$ where $S^{n,n}$ stands for the monoid of $n \times n$ matrices over S with multiplication. Since $\Sigma^\star$ is free, any such homomorphism can be specified by a list of matrices $h(a)$, $a \in \Sigma$.

Exercise

Verify all the examples given above are indeed semirings.

Exercise

Show that all the following structures are semirings.

1. $\langle \mathbb{R} \cup \{+\infty\}, \min, +, \infty, 0 \rangle$,
2. $\langle \mathbb{R} \cup \{-\infty\}, \max, +, -\infty, 0 \rangle$,
3. $\langle \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}, \max, \geq, 0, 1 \rangle$,
4. $\langle \mathbb{R}_{\geq 0} \cup \{+\infty\}, \max, \min, 0, \infty \rangle$.

Exercise (Fusion Semiring)

Recall the partial operation of fusion on words. We have extended fusion to languages as usual (note that fusion on languages is a total operation). Show that the structure $\langle \mathcal{P}(\Gamma^*), \cup, \bullet, \emptyset, \Gamma \rangle$ *is a semiring, the fusion semiring over* Γ.

Star Semirings and Closed Semirings

Consider again the semiring $\mathcal{R}$ of all binary relations on a set A. The multiplicative operation here is composition of relations. As we have seen, there is important operation closely related to composition: transitive closure. How about adding a unary operation * to the semiring that denotes transitive closure?

This additional unary operation also occurs in other semirings. For example, in the semiring $\mathcal{L}(\Sigma)$ of all languages over some alphabet, there is the Kleene star operation, that is of central importance in language theory. As it turns out, the star operation will allow us to solve linear equations of the form $x = a \cdot x + b$.

We can define the star operation in terms of an infinitary operation

$$x^* = \sum_{i \geq 0} x^i = x^0 + x^1 + x^2 + \ldots$$

Note that this approach does not quite fit into the equational framework used to describe the other algebras. Here is a more precise definition. Suppose S is an idempotent semiring with an additional infinitary operation $\sum_{i \in I} a_i$ which is defined for any family $(a_i | i \in I)$ of elements in S where I is an arbitrary index set. S is a closed semiring iff this infinite summation operation behaves properly with respect to finite sums, distributivity and reordering of the arguments. More precisely, we require

$$\begin{aligned} \sum_{i \in [n]} a_i &= a_1 + \ldots + a_n, \\ (\sum_{i \in I} a_i)(\sum_{j \in J} b_j) &= \sum_{(i,j) \in I \times J} a_i b_j, \\ \sum_{i \in I} a_i &= \sum_{j \in J} (\sum_{i \in I_j} a_i), \end{aligned}$$

where in the last equation I is the disjoint union of the sets I_j. The star operation in a closed semiring is then defined as above. We can then verify equations such as

$$(x + y)^* = (x^* y)^* x^*,$$
$$(xy)^* = 1 + x(yx)^* y,$$

It is easy to check that $\mathcal{L}(\Sigma)$ is in fact a closed semiring, since plus here corresponds to set theoretic union and can naturally be expanded to infinitely many arguments. Note that in this case we also have the super idempotency property: $\sum a_i = a$ if for all $i \in I$: $a_i = a$. It follows that $(x^*)^* = x^*$ and $1^* = 1$.

We will denote semirings with a star operation by $\langle X, +, \cdot, ^*, 0, 1 \rangle$. For a more detailed description of closed semirings and star semirings see [**?**, **?**].

Exercise

Show that the collection of all Boolean n *by* n *matrices forms a closed semiring. Think about the case* $n = 1$ *first.*

Exercise

Verify the equations given above in $\mathcal{L}(\Sigma)$. *Solve the equation* $x = a \cdot x + b$ *over* $\mathcal{L}(\{a, b\})$.

Exercise

Show that in a closed semiring with super idempotency we have $(x^*)^* = x^*$ *and* $1^* = 1$. *Also show that these equations hold in any closed semiring satisfying* $(x + y)^* = (x^* y)^* x^*$ *and* $(xy)^* = 1 + x(yx)^* y$.

A ring is a semiring where the additive monoid is an Abelian group. A field is a ring where in addition the multiplicative monoid is Abelian, and is essentially a group: $\langle X - \{0\}, \otimes, 1 \rangle$ is a group.

Note that one cannot give a reasonable definition of 0^{-1}, which is why 0 is excluded from the multiplicative group of a field.

Thus, in a field one has a good chance to solve polynomial equations such as $x^2 - 5x + 3 = 0$. Note, though, that we cannot hope to solve all such equations in all fields: over the real numbers, $x^2 + 1 = 0$ has no solution (but is has a solution over $\mathbb{C}$, the algebraic completion of the reals).

Definition

A ring is an algebraic structure of the form

$$\mathcal{R} = \langle R, +, \cdot, 0, 1 \rangle$$

where

- $\langle R, +, 0 \rangle$ is a commutative group (additive group),
- $\langle R, \cdot, 1 \rangle$ is a monoid (not necessarily commutative),
- multiplication distributes over addition:

$$x \cdot (y + z) = x \cdot y + x \cdot z$$
$$(y + z) \cdot x = y \cdot x + z \cdot x$$

Note that we need two distributive laws since multiplication is not assumed to be commutative. If multiplication is commutative the ring itself is called commutative.

One can relax the conditions a bit and deal with rings without a 1: for example, $2\mathbb{Z}$ is a ring without 1. Instead of a multiplicative monoid one has a semigroup.

For our purposes there is no need for this.

Example (Integers)

The integers $\langle \mathbb{Z}, +, *, 0, 1 \rangle$ with the usual addition and multiplication for a ring.

Example (Modular Numbers)

The integers modulo n, $\langle \mathbb{Z}_n, +, *, 0, 1 \rangle$ with the usual addition and multiplication form a ring. If n is prime, this ring is actually a field. In particular there is a two-element field consisting just of 0 and 1. Note that these fields are finite.

Example (Standard Fields)

The rationals $\mathbb{Q}$, the reals $\mathbb{R}$, the complex numbers $\mathbb{C}$.

Example (Univariate Polynomials)

Given a ring R we can construct a new ring by considering all polynomials with coefficients in R, written $R[x]$ where x indicates the "unknown" or "variable".

For example, $\mathbb{Z}[x]$ is the ring of all polynomials with integer coefficients.

Example (Matrix Rings)

Another important way to construct rings is to consider square matrices with coefficients in a ground ring R.

For example, $\mathbb{R}^{n,n}$ denotes the ring of all n by n matrices with real coefficients. Note that this ring is not commutative unless $n = 1$.

Definition

A ring element a is an annihilator if for all x: $xa = ax = a$.

An inverse u' of a ring element u is any element such that $uu' = u'u = 1$.

A ring element u is called a unit if it has an inverse u'.

Proposition

0 is an annihilator in any ring.

Proof. Note that $a0 = a(0 + 0) = a0 + a0$, done by cancellation in the additive group. □

Inverses

Note that an annihilator cannot be a unit.

For suppose $aa' = 1$. We have $xa = a$, so that $x = 1$, contradiction.

The multiplicative 1 in a ring is uniquely determined: $1 = 1 \cdot 1' = 1'$.

Proposition

If u is a unit, then its inverse is uniquely determined

Proof.

Suppose $uu' = u'u = 1$ and $uu'' = u''u = 1$. Then

$$u' = u'1 = u'uu'' = 1u'' = u''.$$

□

As usual, lots of equational reasoning. And we can write the inverse as u^{-1}.

Integral Domains

We are interested in rings that have lots of units. One obstruction to having a multiplicative inverse is described in the next definition.

Definition

A ring element $a \neq 0$ is a zero divisor if there exist $b, c \neq 0$ such that $ab = ca = 0$.

A commutative ring is an integral domain if it has no zero-divisors.

Consider $R^\star = R - \{0\}$, so all units are located in $R^\star$.

Then $\langle R^\star, \cdot, 1 \rangle$ is a monoid in any integral domain.

Proposition (Multiplicative Cancellation)

In an integral domain we have $ab = ac$ where $a \neq 0$ implies $b = c$.

Proof. $ab = ac$ iff $a(b - c) = 0$, done. □

Examples: Integral Domains

Example (Standard Integral Domains)

The integers $\mathbb{Z}$, the rationals $\mathbb{Q}$, the reals $\mathbb{R}$, the complex numbers $\mathbb{C}$ are all integral domains.

Example (Modular Numbers)

The ring of modular numbers $\mathbb{Z}_m$ is an integral domain iff m is prime.

Example (Non-ID)

The ring of 2×2 real matrices is not an integral domain:

$$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$$

A Strange Ring

Arithmetic structures provide the standard examples for rings, but the axioms are much more general than that. Here is a warning not to over-interpret the ring axioms.

Let A be an arbitrary set and let $P = \mathfrak{P}(A)$ be its powerset. For $x, y \in P$ define

$$x + y = (x - y) \cup (y - x)$$
$$x * y = x \cap y$$

Thus addition is symmetric difference and multiplication is plain set-theoretic intersection. In terms of logic, addition is "exclusive or," and multiplication is "and."

Proposition

$\langle \mathfrak{P}(A), +, *, \emptyset, A \rangle$ *is a commutative ring.*

Exercise

Prove the proposition.

Fields

Definition

A field $\mathbb{F}$ is a ring in which the multiplicative monoid $\langle F^*, \cdot, 1 \rangle$ forms a commutative group.

In other words, every non-zero element is already a unit. As a consequence, in a field we can always solve linear equations

$$a \cdot x + b = 0$$

provided that $a \neq 0$: the solution is $x_0 = -a^{-1}b$. In fact, we can solve systems of linear equations using the standard machinery from linear algebra.

As we will see, this additional condition makes fields much more constrained than arbitrary rings. By the same token, they are also much more manageable.

Examples: Fields

Example

In calculus one always deals with the classical fields: the rationals $\mathbb{Q}$, the reals $\mathbb{R}$, the complex numbers $\mathbb{C}$.

Example

The modular numbers $\mathbb{Z}_m$ form a field for m is prime.

We can use the Extended Euclidean algorithm to compute multiplicative inverses: obtain two cofactors x and y such that $xa + ym = 1$. Then x is the multiplicative inverse of a modulo m.

Note that we can actually compute quite well in this type of finite field: the elements are trivial to implement and there is a reasonably efficient way to realize the field operations.

Axiomatization

Note that one can axiomatize monoids and groups in a purely equational fashion, using a unary function symbol $^{-1}$ to denote an inverse function when necessary.

Alas, this does not work for fields: the inverse operation is partial and we need to guard against argument 0:

$$x \neq 0 \;\Rightarrow\; x * x^{-1} = 1$$

One can try to pretend that inverse is total and explore the corresponding axiomatization; this yields a structure called a "meadow" which does not quite have the right properties.

Products Fail

One standard method in algebra that produces more complicated structures from simpler one is to form a product (operations are performed componentwise).

This works fine for structures with an equational axiomatization: semigroups, monoids, groups, and rings.

Unfortunately, for fields this approach fails. For let

$$F = F_1 \times F_2$$

where F_1 and F_2 are two fields. Then F is a ring, but never a field: the element $(0,1) \in F$ is not $(0,0)$, and so would have to have an inverse (a,b). But $(0,1)(a,b) = (0,b) \neq (1,1)$, so this does not work.

Polynomials

Informally, a (univariate) polynomial is an expression of the form

$$x^3 - 2x^2 + 3x - 1$$

There is an unknown or variable x and we

- form powers x^i of the variable (monomials),
- multiply them by an element in some ground ring (coefficients),
- add several such terms.

Of course, division is not allowed.

The ground ring supplies the coefficients of the polynomials.

In the example, it is presumably $\mathbb{Z}$.

Coefficient Representation

Hence we can represent a polynomial by a vector of its coefficients:

$$\boldsymbol{a} = (a_0, a_1, \ldots, a_{n-1})$$

represents

$$p(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1}$$

If $a_{n-1} \neq 0$ then $n-1$ is the degree of p.

In general, $p(x)$ has degree bound n (note that this is a strict bound; this notion will be useful later in several algorithms).

Evaluation

Suppose we have a univariate polynomial $p(x)$ over ring R.

We can replace the unknown x by elements in R and evaluate to obtain another element in the ring: For example,

$$p(x) = x^3 - 2x^2 + 3x - 1$$

produces

$$p(2) = 2^3 - 2 \cdot 2^2 + 3 \cdot 2 - 1 = 5.$$

Hence we can associate the polynomial p with a polynomial function

$$\widehat{p} : R \to R, \quad a \mapsto p(a)$$

This may seem like splitting hairs, but sometimes it is important to keep the two notions apart.

Come on . . .

The polynomial and the associated polynomial function really are two different objects. Consider the ground ring $\mathbb{Z}_2$. The polynomial

$$p(x) = x + x^2$$

has the associated function

$$\widehat{p}(a) = 0$$

for all $a \in \mathbb{Z}_2$.

In fact, any polynomial $p(x) = \sum_{i \in I} x^i$ produces the identically 0 map as long as $I \subseteq \mathbb{N}^+$ has even cardinality.

Exercise

Describe all polynomial functions over ground rings $\mathbb{Z}_2$ and $\mathbb{Z}_3$.

Horner's Rule

If the polynomial is given in coefficient form

$$\boldsymbol{a} = (a_0, a_1, \ldots, a_d)$$

we can easily evaluate at any point $b \in R$:

$$f(b) = ((\ldots(a_d b + a_{d-1})b + a_{d-2})b + \ldots)b + a_0$$

Proposition

A polynomial of degree d can be evaluated in d ring multiplications and d ring additions.

Interpolation

Suppose we wish to construct a polynomial f that evaluates to given target values at certain points. Say we want $f(a_i) = b_i$ for $i = 0, \ldots, n-1$. Define the Lagrange interpolant

$$L_i^n(x) = \prod_{j \neq i} \frac{x - a_j}{a_i - a_j}$$

Proposition

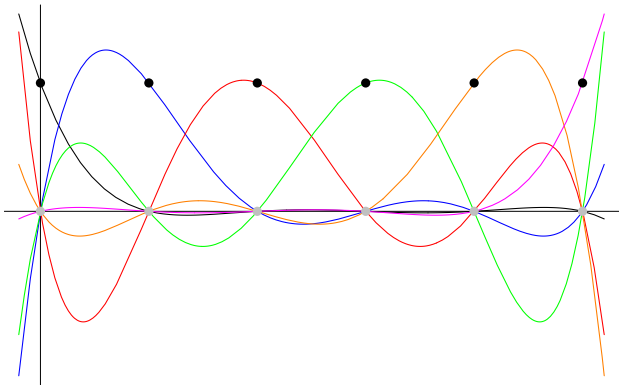
$L_i^n(a_i) = 1$ *and* $L_i^n(a_j) = 0$ *for* $i \neq j$.

Hence we can choose

$$f(x) = \sum_{i<n} b_i L_i^n(x)$$

Note that f has degree bound n.

Lagrange Functions



Example

Suppose we want $f(i) =$ the ith prime for $i = 0, \ldots, 5$.

The Lagrange interpolation looks like

$$f(x) = 2L_0^6 + 3L_1^6 + 5L_2^6 + 7L_3^6 + 11L_4^6 + 13L_5^6$$

which, after expansion and simplification, produces

$$\frac{1}{120}(240 - 286\,x + 735\,x^2 - 425\,x^3 + 105\,x^4 - 9\,x^5)$$

Secret Sharing

Suppose you have a "secret" a, a natural number, that you want to distribute over n people in such a way that no proper subgroup of the n persons can access the secret but the whole group can.

More generally may want to distribute the secret so that k out of n persons can access it, but not any subgroup of size $k-1$.

Here is a simple idea for the $n = k$ problem:

We may safely assume that a is a m-bit number. Generate $n-1$ m-bit numbers a_i and give number a_i to person i, $i = 1, \ldots, n-1$. Person n receives

$$a_n = a \oplus a_1 \oplus a_2 \oplus \ldots \oplus a_{n-1}$$

where $\oplus$ is bit-wise xor.

Clearly all n secret sharers can compute a, but if one is missing they are stuck with a random number.

Using Polynomials

A more flexible approach is built on the following idea.

Construct a polynomial f that has a as its lowest coefficient, so $f(0) = a$. All other coefficients are chosen at random.

The secret sharers are given not the coefficients of f but pieces of the point-value description of f, which are obtained by repeated evaluation.

If all n agree, they can use their information to reconstruct the coefficient representation of f by interpolation.

Once f is reconstructed a can be found be a simple evaluation.

But for any proper subset f will be underdetermined, and a cannot be recovered.

It is not hard to generalize this method to $k < n$.

Shamir's Method

More precisely, pick a prime $p > a, n$. We will use the ground ring $\mathbb{Z}_p$.

Generate random numbers $0 < a_i < p$ for $i = 1, \ldots, n-1$.

Define the polynomial

$$f(x) = a + a_1 x + \ldots + a_{n-1} x^{n-1}$$

f is completely determined by the n point-value pairs (i, b_i), $i = 1, \ldots, n$.

By interpolation we can retrieve f from the point-value pairs, hence we can determine $a = a_0$.

An important point is that $n-1$ persons can obtain no information about the zero coefficient; every coefficient is equally likely.

Defining Polynomials

So polynomials are associated with functions, we can evaluate them efficiently and we can reconstruct them from point-value pairs.

How do we give a precise definition of polynomials?

The definition should pin down all the essential properties and should make the algebraic structure perfectly clear For example, we can add and multiply polynomials.

There are (at least) two ways we could turn:

- Define an algebraic structure $R[x]$, the so-called polynomial ring over R.
- Define a collection of formal expressions representing polynomials (implicit representation).

Pro and Con

The algebraic approach, unsurprisingly, has the advantage that it brings out the algebraic properties of polynomials very clearly. Surprisingly, though, it also provides some ideas for implementation. It is a bit abstract, though.

The formal expression approach is more intuitive and closely follows actual practical use. E.g., an expression like

$$(x + y)^5 - 1$$

will be polynomial in x and y by definition. To evaluate, we perform term substitution followed by some arithmetic evaluations.

Unfortunately, the algebra gets more tricky when dealing with purely syntactic structures.

Polynomial Rings and Coproducts

We start with a moderately strict algebraic definition of polynomial.

Let R be a commutative ring with 1 throughout.

Definition

Given a ring R the $\mathbb{N}$-coproduct of R is defined by

$$\coprod_{\mathbb{N}} R = \{\, (a_i) \in R^{\mathbb{N}} \mid \text{only finitely many } a_i \neq 0 \,\}$$

The sequence notation $(a_i) \in \coprod_{\mathbb{N}} R$ is a bit clumsy since only finitely many terms are non-zero. One usually writes suggestively

$$a_0 + a_1 x + \ldots + a_n x^n$$

where a_n is the last non-zero element in the sequence (or $n = 0$ if they all are 0). Note that the "unknown" x is nothing but syntactic sugar, all we really have is a sequence with finite support.

Operations on Coproducts

Again, there is no "unknown" in the definition of the coproduct. That makes it easier to give clean definitions of the algebraic structure.

Addition is easy:

$$(a_i) + (b_i) = (a_i + b_i)$$

The sum $(a_i) + (b_i)$ is again an element of the coproduct and it is not too hard to check that this operation is associative and commutative.

But multiplication is somewhat more complicated (Cauchy product):

$$(a_i) \cdot (b_i) = \Big(\sum_{r+s=i} a_r \cdot b_s\Big)$$

Proposition

The product $(a_i) \cdot (b_i)$ is an element of the coproduct.

More Arithmetic

Write $\mathbf{0}$ and $\mathbf{1}$ for the sequences $(0, 0, 0, \ldots)$ and $(1, 0, 0, \ldots)$, respectively.

We have $\boldsymbol{a} + \mathbf{0} = \boldsymbol{a}$ so that

$$\left\langle \coprod R, +, \mathbf{0} \right\rangle$$

is a commutative monoid and even a group.

Likewise $\mathbf{1} \cdot \boldsymbol{a} = \boldsymbol{a} \cdot \mathbf{1} = \boldsymbol{a}$ and

$$\left\langle \coprod R, \cdot, \mathbf{1} \right\rangle$$

is also a commutative monoid (assuming that R is commutative).

Exercise

Prove that coproducts are closed under multiplication. Which properties of the natural numbers are important?

Exercise

Prove that $\langle \coprod R, \cdot, \mathbf{1} \rangle$ is a commutative monoid.

The Unknown

Here is a much more interesting element: let

$$x = (0, 1, 0, 0, 0, \ldots)$$

Then $x^2 = (0, 0, 1, 0, 0, \ldots)$, $x^3 = (0, 0, 0, 1, 0, \ldots)$ and so forth.

This justifies the notation

$$a_0 + a_1 x + \ldots + a_n x^n$$

for the sequence

$$(a_0, a_1, \ldots, a_n, 0, 0, \ldots)$$

Note that one can actually embed all of R:

$$r \mapsto (r, 0, 0, \ldots)$$

Moreover, this map is (trivially) a ring monomorphism.

The Polynomial Ring

Lemma

$\langle \coprod R, +, \cdot, \mathbf{0}, \mathbf{1} \rangle$ is a ring. This ring is commutative whenever R is.

This is unsurprising, but note that a proof requires a bit of work: we have to verify e.g. that multiplication as defined above really is associative.

We ignore the details.

Definition

The ring $\langle \coprod R, +, \cdot, \mathbf{0}, \mathbf{1} \rangle$ is the polynomial ring with coefficients in R and is usually written $R[x]$.

In calculus one studies $\mathbb{R}[x]$.

For our purposes, $\mathbb{Q}[x]$, $\mathbb{Z}[x]$, $\mathbb{Z}_m[x]$ or $[x]$ where is a finite field will be more important.

More Abstraction

The last definition may seem fairly abstract, but one can push even further ahead.

Note that $\mathbb{N}$ was not just used to define the coproduct (a sequence is a map $\mathbb{N} \to R$) but that addition on $\mathbb{N}$ was used in the definition of multiplication:

$$(a_i) \cdot (b_i) = \Big(\sum_{r+s=i} a_r \cdot b_s\Big)$$

What we are really using here is not just any old countable set but the monoid

$$\langle \mathbb{N}, +, 0 \rangle$$

The fact that the equation $r + s = i$ has only finitely many solutions in this monoid was crucial.

So? Everybody knows kindergarten arithmetic. Why make a fuss about it?

Generalization

Suppose we have a commutative monoid

$$\langle M, +, 0 \rangle$$

where equations $x + y = m$ have only finitely many solutions.

Then we can define a ring $R[M]$ on the carrier set

$$\coprod_M R = \{ \boldsymbol{a} : M \to R \mid \text{only finitely many non-zeros } \}$$

in the exact same way as before.

We need at least one example for M (other then $\mathbb{N}$) that makes this construction interesting. It is probably good to stay close to $\mathbb{N}$. $\mathbb{Z}$ won't work since it violates the finiteness conditions.

How about $M = \mathbb{N} \times \mathbb{N}$?

Bivariate Polynomials

Instead of sequences we now have a grid of coefficients (a two-dimensional array instead of a one-dimensional one).

$$\begin{pmatrix} a_{00} & a_{01} & a_{02} & \dots & a_{0j} & \dots \\ a_{10} & a_{11} & a_{12} & \dots & a_{1j} & \dots \\ a_{20} & a_{21} & a_{22} & \dots & a_{2j} & \dots \\ & \vdots & & \ddots & & \\ a_{i0} & a_{i1} & a_{i2} & \dots & a_{ij} & \dots \\ & \vdots & & & & \end{pmatrix}$$

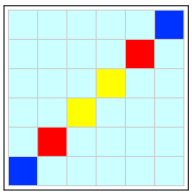
Addition in $R[\mathbb{N} \times \mathbb{N}]$ is essentially the same as before (since it does not depend on the algebraic structure of M, just the carrier set).

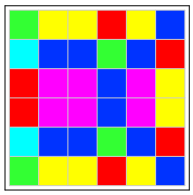
But multiplication becomes more interesting.

Multiplication

Let $x = \begin{pmatrix} 0&0&0&\dots \\ 1&0&0&\dots \\ 0&0&0&\dots \\ \vdots&\vdots&\vdots&\ddots \end{pmatrix}$ and $y = \begin{pmatrix} 0&1&0&\dots \\ 0&0&0&\dots \\ 0&0&0&\dots \\ \vdots&\vdots&\vdots&\ddots \end{pmatrix}$.

Then for $R = \mathbb{Z}_7$ the products $(x + y)^5$ and $(1 + x)^5(2 + y)^5$ look like this (all other entries are 0).





Multivariate Polynomials

One can easily show that any element in $R[\mathbb{N} \times \mathbb{N}]$ has the form

$$\boldsymbol{a} = \sum_{i,j \geq 0} a_{i,j} x^i y^j$$

where only finitely many coefficients are non-zero and addition and multiplication work as one would expect. So we really have a rock-solid definition of bivariate polynomials. And, of course, using the monoid

$$M = \langle \mathbb{N}^n, +, 0 \rangle$$

we can get multivariate polynomials in general.

Definition

The ring of polynomials over R in n variables is defined by

$$R[x_1, \dots, x_n] = R[\mathbb{N}^n]$$

Normal Form

There is a natural way to write down a multivariate polynomial analogous to the univariate and bivariate case. A monomial is a term of the form

$$\boldsymbol{x}^{\boldsymbol{e}} = x_1^{e_1} x_2^{e_2} \dots x_n^{e_n}.$$

where $\boldsymbol{e} = (e_1, \dots, e_n) \in \mathbb{N}^n$. Then any multivariate polynomial can be written as a sum of monomials, multiplied by the appropriate coefficients.

$$p(\boldsymbol{x}) = \sum a_{\boldsymbol{e}}\, \boldsymbol{x}^{\boldsymbol{e}}$$

And, arithmetic works as expected.

Note that the sum is naturally ordered (use the natural lex order on the exponents).

Note: Some authors consider the coefficient to be part of the monomial.

Hold It

Why not simply define, say, bivariate polynomials by saying they are expressions like

$$p(x, y) = x^2 y^2 + 3x - y + 1$$

Or, if you want more precision, since the operation $R \mapsto R[x]$ works for any ring R we can simply repeat it to get multivariate polynomials:

$$R[x_1, \dots, x_n] = R[x_1, \dots, x_{n-1}][x_n]$$

All true, but these "definitions" obscure a lot of things.

- What is $R[x][x]$, what $R[x, x]$?
- What is the difference between $R[x]$, $R[X]$ and $R[y]$?
- What is the difference between $R[x, y]$, $R[y, x]$?
- What is the role of evaluation?

Clarity

One might argue that $R[x][x]$ and $R[x,x]$ simply make no sense.

$R[x]$ and $R[y]$ are isomorphic, so in a sense they are the same.

For a human being (a mathematician) these difficulties are usually irrelevant: one can determine from context what is meant, or even correct the author if need be.

But if one wants to implement polynomial algebra things are more problematic. For example, if evaluation is done by rewrite rules, the name of the variable does play a huge role. Here $R[x]$ and $R[y]$ can behave very differently.

Also note that from the implementation point of view $\coprod_{\mathbb{N}^n} R$ is probably a better model than definitions using terms.

Evaluation

So what exactly is the role of evaluation?

The point is that once we fix the value of the unknown(s), everything else is determined, too.

More precisely, suppose $f : R \to S$ is some ring homomorphism. Consider the set of all ring homomorphisms from $R[x]$ to S that agree with f on R:

$$H = \{\, h : R[x] \to S \mid h \restriction R = f \,\}$$

Then the map

$$H \to S \qquad h \mapsto h(x)$$

is a bijection.

Universal Property

Another way of saying the same thing:

For any homomorphism $f : R \to S$ and any element $a \in S$ there is exactly one homomorphism $h : R[x] \to S$ such that

- $h \restriction R = f$ and
- $h(x) = a$.

Exercise

Give a similar result for multivariate polynomials.

Exercise

Show that this property in fact characterizes polynomial rings.

Degrees

Definition

The degree of a monomial $\boldsymbol{x}^{\boldsymbol{e}}$ is $\sum e_i$. The degree of a polynomial is the largest degree of any of its monomials. If the polynomial is zero its degree is $-\infty$.

Lemma

Let R be an integral domain and $p, q \in R[x]$. Then $\deg(pq) = \deg(p) + \deg(q)$.

Note that if R be an integral domain then by the lemma $R[x]$ is again an integral domain.

The lemma fails without the integral domain assumption: Let $R = \mathbb{Z}_4$ and consider $p(x) = q(x) = 2x$.

Roots

Definition

A ring element $a \in R$ is a root of $p(x) \in R[x]$ if $p(a) = 0$.

In other words, a root is any solution of the equation $p(x) = 0$.

Finding roots of polynomial equations is often very difficult, in particular when several variables are involved.

For univariate polynomials over the reals good numerical methods exist, but over other rings things are problematic.

For example, computing square roots, i.e. solving $x^2 - a = 0$, over $\mathbb{Z}_m$ is surprisingly difficult. Of course there is a brute-force algorithm, but think of modulus m having thousands of digits.

And for $\mathbb{Z}[x_1, x_2, \ldots, x_n]$ it is even undecidable whether a root exists.

Polynomials and Division

Definition

Given two polynomials f and g, g divides f if for some polynomial q: $q \cdot g = f$.

For the integers, the most important algorithm associated with the notion of divisibility is the Division Algorithm: we can compute quotient q and remainder r such that $a = qb + r$, $0 \leq r < b$.

The situation for polynomials is very similar.

Theorem (Division Algorithm)

Assume that F is a field. Let f and g be two univariate polynomials over F, $g \neq 0$. Then there exist polynomials q and r such that

$$f = q \cdot g + r \qquad \textit{where } \deg(r) < \deg(g).$$

Moreover, q and r are uniquely determined.

Proof Sketch

For existence consider the set of possible remainders

$$S = \{ f - qg \mid q \in F[x] \}.$$

If $\mathbf{0} \in S$ we are done, so suppose otherwise.

Trick: let $r \in S$ be any element of minimal degree, say $r = f - qg$.

Write $m = \deg(r)$ and $n = \deg(g)$, so we need $m < n$.

Assume $m \geq n$ and define

$$r' = r - a_m / b_n \, x^{m-n} \, g$$

where a_m and b_n are the leading coefficients of r and g, respectively.

But then $\deg(r') < \deg(r)$ and $r' \in S$, contradicting minimality.

Uniqueness is left as an exercise.

The Actual Algorithm

Though the theorem is often referred to as "Division Algorithm" it's just an existence and uniqueness result. However, with a little work one can turn the proof into an algorithm.

Write (h) for the leading coefficient of a polynomial h. Suppose g is monic, so that $(g) = 1$.

Here is an abstract version, to be called on f.

remainder(f, g)

r = f; while(deg(r) ¿= deg(g)) c = lc(r); k = deg(r) - deg(g); r = r - c * x$^k * g; returnr;$

Another Version

Often one needs to compute both q and r, here is an array-based version which does that. Assume $\deg(f) = n$ and $\deg(g) = m$.

r = f;

for i = n-m downto 0 do if(deg r = m + i) q[i] = lc(r); r = r - q[i] * x$^i * g; elseq[i] = 0;$

return q[];

Note that sparseness is a tricky issue here: divide $x^{n+1} - 1$ by $x - 1$.

Application: GCD

An important application of the Division Algorithm for integers is the Euclidean algorithm for the GCD.

Likewise we can obtain a polynomial GCD algorithm from the Division Algorithm for polynomials.

In fact, essentially the same algorithm works, just replace $\mathbb{Z}$ by $\mathbb{Z}[x]$.

For example, we can obtain cofactors s and t such that

$$\gcd(f, g) = sf + tg.$$

Local Views

Consider the polynomial

$$p(x) = x^4 - 8x^3 + 23x^2 - 28x + 12$$

We can read off immediately that $p(0) = 12$ and that the tangent at that point has slope -28.

But what about the polynomial around $x = 1$? Here it is convenient to use Taylor expansion to rewrite p in the form

$$p(x) = (x-1)^4 - 4(x-1)^3 + 5(x-1)^2 - 2(x-1)$$

We can see that $x = 1$ is a root and the tangent has slope -2.

More Local Views

Similarly

$$p(x) = (x-2)^4 - (x-2)^2$$
$$p(x) = (x-3)^4 + 4(x-3)^3 + 5(x-3)^2 + 2(x-3)$$

so it is clearly useful to consider polynomials in non-expanded form.

Exercise

What conclusions can you draw from the various representations of $p(x)$?

Bounds

On rare occasions one can also easily establish bounds given the "right" representation of a polynomial. For instance,

$$p(x) = x^4 - 4x^3 + 7x^2 - 10x + 10$$

appears to be positive over $\mathbb{R}$.

One can prove this by tediously checking the behavior over the intervals with endpoints $-\infty$, -1, 1, $3/2$, 2, ∞. Alternatively, one can use differentiation to find the global minimum of p.

Or one can note that $p(x) = (x^2 - 1)^2 + (x - 3)^2$.

Roots and Divisibility

Lemma

Let a be a root of $f \in F[x]$. Then $(x - a)$ divides $f(x)$.

Proof.

Write

$$f = q(x - a) + r$$

where $\deg(r) < 1$. But then r must be 0, done.

□

Lemma

Any non-zero polynomial $f \in F[x]$ has at most $\deg(f)$ many roots.

Proof.

Use the last lemma and induction on the degree.

□

Decomposition

So if $\deg(f) = n$ and f has n roots we decompose f completely into linear terms:

$$f = c(x - a_1)(x - a_2)\dots(x - a_n)$$

Of course, there may be fewer roots, even over a rich field such as $\mathbb{R}$: $f = x^2 + 2$ has no roots.

This problem can be fixed by enlarging $\mathbb{R}$ to the field of complex numbers $\mathbb{C}$ (the so-called algebraic completion of $\mathbb{R}$).

More Roots

Note that over arbitrary rings more roots may well exist.

For example over $R = \mathbb{Z}_{15}$ the equation $x^2 - 4 = 0$ has four roots: $\{2, 7, 8, 13\}$.

But of course

$$(x - 2)(x - 7)(x - 8)(x - 13) = 1 + 7x^2 + x^4 \not\equiv x^2 - 4$$

Exercise

Using the Chinese Remainder theorem explain why there are four roots in the example above. Can you generalize?

Point-Value Representation

The fact that a non-zero polynomial of degree n can have at most n roots can be used to show that the interpolating polynomial

$$f(x) = \sum_i b_i \prod_{j \neq i} \frac{x - a_j}{a_i - a_j}$$

is unique: suppose g is another interpolating polynomial so that $g(a_i) = b_i$.

Then $f - g$ has $n + 1$ roots and so is identically zero.

Hence we have an alternative representation for polynomials: we can give a list of point-value pairs rather than a list of coefficients.

To the naked eye this proposal may seem absurd: why bother with a representation that is clearly more complicated? As we well see, there are occasions when point-value is computationally superior to coefficient list.

Multiplying Polynomials

Suppose we have two univariate polynomials f and g of degree bound n.

Using the brute force algorithm (i.e., literally implementing the definition of multiplication in $\prod R$) we can compute the product fg in $\Theta(n^2)$ ring operations.

Now suppose we are dealing with real polynomials. There is a bizarre way to speed up multiplication:

- Convert f and g into point-value representation where the support points are carefully chosen.
- Multiply the values pointwise to get h.
- Convert h back to coefficient representation.

FFT

It may seem absurd to spend all the effort to convert between coefficient representation and point-value representation. Surprisingly, it turns out that the conversions can be handled in $\Theta(n \log n)$ steps using a technique called Fast Fourier Transform.

But the pointwise multiplication is linear in n, so the whole algorithm is just $\Theta(n \log n)$.

Theorem

Two real polynomials of degree bound n can be multiplied in $\Theta(n \log n)$ steps.

Take a look at CLR for details.

Vandermonde Matrices

Here is another look at conversions between coefficient and point-value representation, i.e., between evaluation and interpolating.

Definition

Define the n by n Vandermonde matrix by

$$(x_0, x_1, \ldots, x_{n-1}) = \begin{pmatrix} 1 & x_0 & x_0^2 & \ldots & x_0^{n-1} \\ 1 & x_1 & x_1^2 & \ldots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n-1} & x_{n-1}^2 & \ldots & x_{n-1}^{n-1} \end{pmatrix}$$

Lemma

$$|(\boldsymbol{x})| = \prod_{i<j} x_j - x_i$$

Evaluation versus Interpolation

It follows that the Vandermonde matrix is invertible iff all the x_i are distinct. Now consider a polynomial

$$f(x) = c_0 + c_1 x + \ldots + c_{n-1} x^{n-1}$$

To evaluate f at points $\boldsymbol{a} = (a_0, \ldots, a_{n-1})$ we can use matrix-by-vector multiplication:

$$\boldsymbol{b} = (\boldsymbol{a}) \cdot \boldsymbol{c}$$

But given the values $\boldsymbol{b}$ we can obtain the coefficient vector by

$$\boldsymbol{c} = (\boldsymbol{a})^{-1} \cdot \boldsymbol{b}$$

Implicit Representation

Recall our alternative way of defining polynomials: formal expressions involving ring elements and the variables using operations plus and times.

We have used this implicit representation in several places already, without any protest from the audience.

- The expressions $c(x - a_1)(x - a_2) \ldots (x - a_n)$ encountered in the root decomposition.
- The description of the determinant of a Vandermonde matrix $\prod_{i<j}(x_j - x_i)$.
- Using interpolation and evaluation to retrieve the secret in Shamir's method.

So?

More precisely, we could consider polynomials to be arbitrary expressions built from

- the variables $x_1, \ldots, x_n$,
- elements in the ground ring,
- addition, subtraction and multiplication.

Obviously we can recover the explicit polynomial (i.e. the coefficient list) from these explicit representations.

Example

Polynomial $p = (x_1 - x_2)(x_3 - x_4)(x_5 - x_6)$ expands to

$$x_1 x_3 x_5 - x_2 x_3 x_5 - x_1 x_4 x_5 + x_2 x_4 x_5 - x_1 x_3 x_6 + x_2 x_3 x_6 + x_1 x_4 x_6 - x_2 x_4 x_6.$$

We just have to expand (multiply out) to get the "classical form".

What exactly is meant by "expanding" a polynomial?

Expanding Polynomials

We want to bring a multivariate polynomial $f(x_1, x_2, \ldots, x_n)$ into normal form. First we apply rewrite rules to push multiplication to the bottom of the tree until we have a sum of products:

- $\alpha(\beta + \gamma) \mapsto \alpha\beta + \alpha\gamma$
- $(\beta + \gamma)\alpha \mapsto \beta\alpha + \gamma\alpha$

Then we collect terms with the same monomial and adjust the coefficient.

- $\ldots + c\boldsymbol{x}^{\boldsymbol{e}} + \ldots + d\boldsymbol{x}^{\boldsymbol{e}} \ldots \mapsto \ldots + (c + d)\boldsymbol{x}^{\boldsymbol{e}} + \ldots$

Some terms may cancel here (we don't keep monomials with coefficient 0).

Computationally it is probably best to first sort the terms (rather than trying to do pattern matching at a distance).

The problem is that it may take exponential time to perform the expansion: there may be exponentially many terms in the actual polynomial.

Polynomial Identities

A task one encounters frequently in symbolic computation is to check whether two polynomials are equivalent, i.e., whether an equation between polynomials

$$f(x_1, x_2, \ldots, x_n) = g(x_1, x_2, \ldots, x_n)$$

holds in the sense that for all $x_1, x_2, \ldots, x_n \in R$.

Definition

Two polynomials $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are equivalent if for all $\boldsymbol{a} \in R$: $f(\boldsymbol{a}) = g(\boldsymbol{a})$. In particular f is identically zero if $f(\boldsymbol{x})$ and 0 are equivalent.

In other words, two polynomials $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are equivalent iff $\widehat{f}(\boldsymbol{x}) = \widehat{g}(\boldsymbol{x})$.

Notation: $f(\boldsymbol{x}) \equiv g(\boldsymbol{x})$.

Zero Testing

Note that the polynomial identity $f \equiv g$ can be rewritten as $f - g \equiv 0$.

Problem:

How can we check whether a multivariate polynomial $f \in F[\boldsymbol{x}]$ is identically zero?

Note: The polynomial may not be given in normal form, but as in the example in a much shorter, parenthesized form. We want a method that is reasonably fast without having to expand out the polynomial first.

Using Roots

Assume that the ground ring is a field F.

Suppose f has degree d. As we have seen, in the univariate case there are at most d roots and d roots determine the polynomial (except for a multiplicative constant).

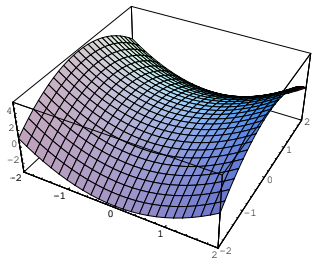
So we could simply check if the polynomial vanishes at, say, $a = 0, 1, 2, \ldots, d$: f will vanish on all $d + 1$ points iff it is identically zero.

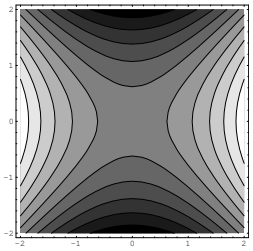
Requires only $d + 1$ evaluations of the polynomial (in any form).

Unfortunately, the roots for multivariate polynomials are a bit more complicated.

A Quadratic Surface

$x^2 - y^2 + 1$





The Key lemma

Lemma

Let $f \in F[x_1, \ldots, x_n]$ be of degree d and $S \subseteq F$ a set of cardinality s. Then f is not identically zero then f has at most ds^{n-1} roots in S^n.

Proof.

The proof is by induction on n.

□

The set S here could be anything. For example, over $\mathbb{Q}$ we might choose $S = \{0, 1, 2, \ldots, s - 1\}$.

Schwartz's Method

The main application of the lemma is to give a probabilistic algorithm to check whether a polynomial is zero.

Suppose f is not identically zero and has degree d.

Choose a point $\boldsymbol{a} \in S^n$ uniformly at random and evaluate $f(\boldsymbol{a})$. Then

$$\Pr[f(\boldsymbol{a}) = 0] \leq \frac{d}{s}$$

So by selecting S of cardinality $2d$ the error probability is $1/2$.

Note that the number of variables plays no role in the error bound.

A Real Algorithm

To lower the error bound, we repeat the basic step k times.

// is f identically zero ?

do k times pick a uniformly at random in S$^n if(f(a)! = 0) return false; od$

return true;

Note that the answer `false` is always correct.

Answer `true` is correct with error bound ε provided that

$$k = \lceil \log 1/\varepsilon \rceil$$

Certainty

With more work we can make sure the f really vanishes.

Corollary

Let $f \in F[x_1, \ldots, x_n]$ be of degree d and $S \subseteq F$ a set of cardinality $s > d$. If f vanishes on S^n then f is identically zero.

But note that for finite fields we may not be able to select a set of cardinality higher than d.

Recall the example over $\mathbb{Z}_2$: in this case we can essentially only choose $S = \mathbb{Z}_2$, so only degree 1 polynomials can be tackled by the lemma.

That's fine for univariate polynomials (since any monomial x^i simplifies to x) but useless for multivariate polynomials.

Application

Recall that a perfect matching in a bipartite graph is a subset M of the edges such that the edges do not overlap and every vertex is incident upon one edge in M.

There is a polynomial time algorithm to check whether a perfect matching exists, but using Schwartz's lemma one obtains a faster and less complicated algorithm.

Suppose the vertices are partitioned into u_i and v_i, $i = 1, \ldots, n$.

Define a $n \times n$ matrix A by

$$A(i,j) = \begin{cases} x_{ij} & \text{if } (u_i, v_j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Note that the determinant of A is a polynomial whose non-zero terms look like

$$\pm x_{1\pi(1)} x_{2\pi(2)} \ldots x_{n\pi(n)}$$

Testing Perfect Matchings

Proposition

The graph has a perfect matching iff the determinant of A is not identically zero.

Proof.

If the graph has no perfect matching then all the terms in the determinant are 0 (they all involve at least one non-edge).

But if the graph has a perfect matching it must have the form $M = \{ (u_i, v_{\pi(i)}) \mid i \in [n] \}$ where π is a permutation.

But then the determinant cannot be 0 since the corresponding monomial cannot be canceled out.

□