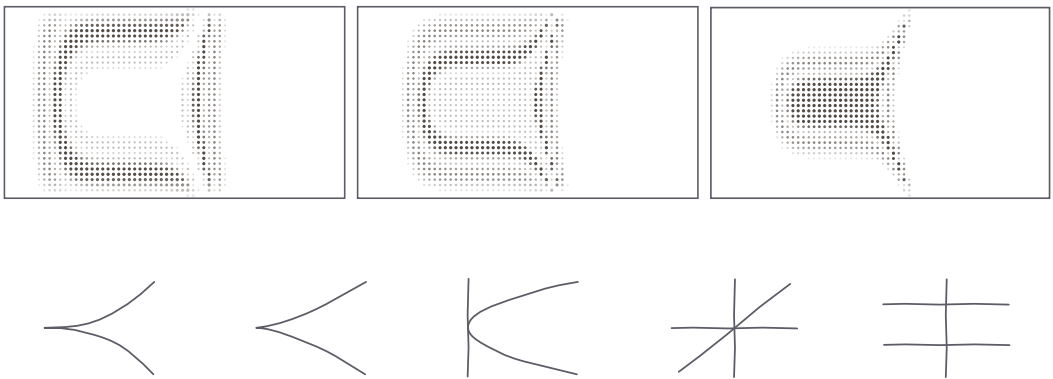
Algebraic Geometry Fundamentals and Resolution of Singularities

Singular Learning Theory, Lecture 2

Aurelio Carlucci

10 November 2026

Department of Computer Science, University of Oxford



Pictures: Matthew Farrugia-Roberts. Ornaments after Franklin Vera Pacheco, *Functorial Resolution of Singularities* (Wikimedia Commons), CC BY-SA 3.0.

The plan

First half

1. The problem
2. Vanishing loci and ideals
3. The Nullstellensatz and schemes
4. Non-reduced schemes: the double point
5. Tangent spaces and singular points
6. Curves and divisors

Second half

7. Simple normal crossings
8. The projective line
9. The blow-up of the plane
10. An example: the cusp
11. Resolution of singularities
12. Summary

1

The problem

Lecture 1: minima with the same Hessian

In Lecture 1: $L_{k,l}(a,b) = a^{2k} + b^{2l}$ at the origin, with $l = 1$.

k	$L_{k,1}(a,b)$	along a	$H(0)$	minimum
1	$a^2 + b^2$	quadratic	$\left(\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right)$	regular
0	$1 + b^2$	flat	$\left(\begin{smallmatrix} 0 & 0 \\ 0 & 2 \end{smallmatrix}\right)$	degenerate
2	$a^4 + b^2$	quartic	$\left(\begin{smallmatrix} 0 & 0 \\ 0 & 2 \end{smallmatrix}\right)$	degenerate
3	$a^6 + b^2$	sextic	$\left(\begin{smallmatrix} 0 & 0 \\ 0 & 2 \end{smallmatrix}\right)$	degenerate

The Hessian depends only on the second-order terms, so it is the same in the last three rows.

Three models with the same zero set

Lecture 1's setting: a parameter–function map $\Phi : \mathcal{W} \to \mathcal{F}$, $w \mapsto f_w$; the shifted loss $K(w) = L(w) - L(w_0) \geq 0$; the *true parameters* $W_0 = \{K = 0\}$. Here: inputs with $\mathbb{E}[x^2] = 1$ and teacher $f_\star = 0$, so that $K(w) = \frac{1}{2}\,\mathbb{E}\big[f_w(x)^2\big]$.

$f_w(x)$	K	true parameters	Hessian at 0
$w\,x$	$\frac{1}{2}w^2$	$w = 0$: the point 0	1
w^2x	$\frac{1}{2}w^4$	$w^2 = 0$: the point 0	0
w^3x	$\frac{1}{2}w^6$	$w^3 = 0$: the point 0	0

All three have the zero set $\{0\}$, and the last two have the same Hessian. Today: an object that tells w^2x from w^3x. Lecture 3 attaches to each model a number that tells all three apart. Every model today is polynomial in w; when that is enough for SLT comes at the end.

2

Vanishing loci and ideals

Affine space, vanishing loci and ideals

Definition: affine space $\mathbb{A}^d_{\Bbbk}$, or $\mathbb{A}^d$

$\mathbb{A}^d_{\Bbbk}$ is $\Bbbk^d$ as a collection of points, not a vector space. Here $\Bbbk = \mathbb{R}$ or $\mathbb{C}$.

Definition: vanishing locus $\mathbb{V}(F)$ (for one function: zero set)

$\mathbb{V}(F) = \{\, P \in \mathbb{A}^d : f(P) = 0 \ \ \forall f \in F \,\}$, for F a set of polynomials.

Definition: ideal I; the ideal $\langle F \rangle$ generated by F

$f, g \in I \implies f + g \in I$; $\ f \in I \implies hf \in I$ for any polynomial h.
$\langle F \rangle = \{\, \sum_i g_i f_i : f_i \in F$ and g_i polynomials, finitely many non-zero $\,\}$;
$(f_1, \ldots, f_N) = \langle \{f_1, \ldots, f_N\} \rangle$.

$\mathbb{V}(F) = \mathbb{V}(\langle F \rangle)$, so we may take F to be an ideal, I.

Four curves in the plane

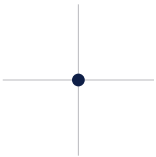
$\mathbb{V}(b)$: a line



$\mathbb{V}(ab)$:
the node



$\mathbb{V}(b^2 - a^3)$:
the cusp



$\mathbb{V}_{\mathbb{R}}(a^2 + b^2)$:
a point

Coordinates (a, b). $\mathbb{V}_{\mathbb{R}}$ is the set of real points. Over $\mathbb{C}$, $a^2 + b^2 = 0$ is two lines, $a = \pm ib$.

How vanishing loci combine

Proposition

For ideals I, J and a family (I_α) of ideals,

$$\mathbb{V}(I) \cup \mathbb{V}(J) = \mathbb{V}(IJ) = \mathbb{V}(I \cap J), \qquad \bigcap_\alpha \mathbb{V}(I_\alpha) = \mathbb{V}\Big(\sum_\alpha I_\alpha\Big),$$

with $IJ = \langle fg : f \in I,\ g \in J \rangle$ and $\sum_\alpha I_\alpha = \big\langle \bigcup_\alpha I_\alpha \big\rangle$. Also $\mathbb{V}(0) = \mathbb{A}^d$, $\mathbb{V}(1) = \emptyset$.

Warning: $IJ \subseteq I \cap J$, but $IJ \neq I \cap J$ in general

For $I = J = (w)$: $(w)(w) = (w^2)$, but $(w) \cap (w) = (w)$; yet $\mathbb{V}(w^2) = \mathbb{V}(w) = \{0\}$. Different ideals with the same zero set.

Irreducible components

Definition: irreducible vanishing locus Z

$Z \neq \emptyset$; $Z = Z_1 \cup Z_2$ with Z_1, Z_2 vanishing loci $\implies$ $Z = Z_1$ or $Z = Z_2$.

Definition: irreducible components of Z

$Z = Z_1 \cup \cdots \cup Z_r$ with the Z_i irreducible, none inside another: exists, unique up to order.

- $\mathbb{V}(ab) = \mathbb{V}(a) \cup \mathbb{V}(b)$: two components. That is the case for the model $ab\,x$, with $K = \frac{1}{2}a^2b^2$.
- $f = c\,g_1^{\ell_1} \cdots g_r^{\ell_r}$ with the g_i distinct and irreducible: over $\mathbb{C}$, the components of $\mathbb{V}(f)$ are the $\mathbb{V}(g_i)$. The field matters: $a^2 + b^2$ is irreducible over $\mathbb{R}$, but equals $(a + ib)(a - ib)$ over $\mathbb{C}$.
- The vanishing loci are, by definition, the closed sets of the *Zariski topology*.

3

The Nullstellensatz and schemes

The ideal of a set

Definition: ideal $\mathbb{I}(Z)$ of a set $Z \subseteq \mathbb{A}^d$

$$\mathbb{I}(Z) = \{\, f \in \Bbbk[w_1, \ldots, w_d] : f(P) = 0 \ \ \forall P \in Z \,\}, \quad \text{an ideal.}$$

- Both $\mathbb{V}$ and $\mathbb{I}$ reverse inclusions.
- $F \subseteq \mathbb{I}\big(\mathbb{V}(F)\big)$ and $Z \subseteq \mathbb{V}\big(\mathbb{I}(Z)\big)$, with equality in the second exactly when Z is a vanishing locus.
- How far the first is from an equality: the Nullstellensatz.

The Nullstellensatz

Theorem (Hilbert)

Let $\Bbbk = \mathbb{C}$, and let J be an ideal. Then

$$\mathbb{I}\big(\mathbb{V}(J)\big) = \sqrt{J} = \{\, f : f^N \in J \text{ for some } N \geq 1 \,\},$$

the *radical* of J.

So $J \mapsto \mathbb{V}(J)$ is one to one on *radical* ideals, those with $\sqrt{J} = J$.

Over $\mathbb{R}$ it fails

$\mathbb{V}_{\mathbb{R}}(a^2 + b^2) = \{0\}$, so $\mathbb{I}\big(\mathbb{V}_{\mathbb{R}}(a^2 + b^2)\big) = (a, b)$, while $\sqrt{(a^2 + b^2)} = (a^2 + b^2)$ in $\mathbb{R}[a, b]$.

The real Nullstellensatz

Theorem (Krivine 1964; Dubois 1970; Risler 1970)

Over $\mathbb{R}$, for an ideal J,

$$\mathbb{I}\big(\mathbb{V}_{\mathbb{R}}(J)\big) = \sqrt[\mathbb{R}]{J},$$

where the *real radical* $\sqrt[\mathbb{R}]{J}$ is the set of polynomials f with $f^{2N} + g_1^2 + \cdots + g_s^2 \in J$ for some $N \geq 1$ and polynomials $g_1, \ldots, g_s$. (See also Bochnak, Coste and Roy 1998.)

Example: $a^2 + b^2$ over $\mathbb{R}$

$a^2 + b^2 \in J$ puts a and b in $\sqrt[\mathbb{R}]{J}$ (take $N = 1$ and one g), so $\sqrt[\mathbb{R}]{(a^2 + b^2)} = (a, b)$ is the ideal of the origin.

Different ideals with the same radical

In one variable

$$\mathbb{V}(w) = \mathbb{V}(w^2) = \mathbb{V}(w^3) = \{0\}, \qquad \sqrt{(w)} = \sqrt{(w^2)} = \sqrt{(w^3)} = (w).$$

Key point

The radical forgets the order of vanishing. The ideals (w), (w^2), (w^3) are different; their zero sets are the same.

The three models' true parameters satisfy $w = 0$, $w^2 = 0$ and $w^3 = 0$: three ideals, (w), (w^2) and (w^3), with one radical, (w).

Functions on a set

Let $Z \subseteq \mathbb{A}^d$ be any subset of affine space. Polynomials f and g agree on Z exactly when $f - g \in \mathbb{I}(Z)$. So the functions on Z form the *quotient ring* $\Bbbk[Z] = \Bbbk[w_1, \ldots, w_d] \,/\, \mathbb{I}(Z)$: polynomials, two of them identified when their difference lies in $\mathbb{I}(Z)$.

Z	$\mathbb{I}(Z)$	$\Bbbk[Z]$
a point	$(w-3)$	$\Bbbk[w]/(w-3) \cong \Bbbk, \quad f \mapsto f(3)$
a line	$(a-b+1)$	$\cong \Bbbk[a]$, with $b = a+1$
the circle	(a^2+b^2-1)	$\Bbbk[a,b] \,/\, (a^2+b^2-1)$

$\mathbb{I}(Z)$ is radical, so these rings have no non-zero *nilpotents* (no $f \neq 0$ with $f^N = 0$): they are *reduced*.

The scheme of an ideal

Allow the ring $\Bbbk[w_1, \ldots, w_d]/J$ for *any* ideal J, radical or not.

The scheme $X = (J = 0)$ with equations J

- $(J_1 = 0) = (J_2 = 0)$ exactly when $J_1 = J_2$.
- Its *support* is the set $\{J = 0\} = \mathbb{V}(J)$.
- Its ring of functions is $\mathcal{O}_X = \Bbbk[w_1, \ldots, w_d]/J$.

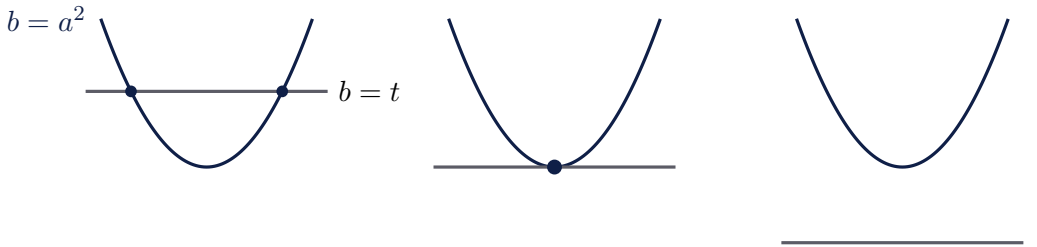
We also write $X : J = 0$; for $J = (f_1, \ldots, f_N)$, also $(f_1 = \cdots = f_N = 0)$, as in $(w^2 = 0)$.

For $J = \mathbb{I}(Z)$, with Z a set: $\mathcal{O}_X = \Bbbk[Z]$, nothing new. For $J = (w^2)$: $\mathcal{O}_X = \Bbbk[w]/(w^2)$, while $\Bbbk[\{0\}] = \Bbbk$: the double point.

4

Non-reduced schemes: the double point

Two points become a double point



$t > 0$: two points $\qquad$ $t = 0$: one double point $\qquad$ $t < 0$: no real points

$\mathbb{k}[a, b] \,/\, (b - a^2,\ b - t) \ \cong\ \mathbb{k}[a] \,/\, (a^2 - t)$: dimension 2 over $\mathbb{k}$, for every t.

The double point $\mathbb{D}$

Definition: dual numbers $\Bbbk[\varepsilon]$

$\Bbbk[\varepsilon] = \Bbbk[w]/(w^2)$, where ε is the class of w: $\varepsilon \neq 0$ and $\varepsilon^2 = 0$. Elements $s + t\varepsilon$, with $(s_1 + t_1\varepsilon)(s_2 + t_2\varepsilon) = s_1 s_2 + (s_1 t_2 + s_2 t_1)\,\varepsilon$.

Definition: double point $\mathbb{D}$

$\mathbb{D} = (w^2 = 0)$, with ring $\Bbbk[\varepsilon] = \mathcal{O}_{\mathbb{D}}$. Its support is the single point 0, and a function f on $\mathbb{D}$ is determined by $f(0)$ and $f'(0)$.

- For a polynomial f, $f(s+\varepsilon) = f(s) + f'(s)\,\varepsilon$. Multiplying two such expressions in $\Bbbk[\varepsilon]$ gives the product rule, $(fg)' = f'g + fg'$.
- ε is a non-zero nilpotent, so $\Bbbk[\varepsilon]$ is not reduced. At $t = 0$, the ring $\Bbbk[a]/(a^2 - t)$ is $\Bbbk[\varepsilon]$: the parabola and its tangent line meet in a non-reduced scheme.

Fat points and the three models

The three models' equations give three schemes:

model	scheme	ring of functions	dim over $\mathbb{k}$	support
$w\,x$	$(w=0)$	$\mathbb{k}$	1	$\{0\}$
w^2x	$(w^2=0)$	$\mathbb{k}[w]/(w^2)=\mathbb{k}[\varepsilon]$	2	$\{0\}$
w^3x	$(w^3=0)$	$\mathbb{k}[w]/(w^3)$	3	$\{0\}$

Three different schemes with the same support. The ring tells w^2x from w^3x, where the Hessian could not.

In general, the *fat point* $(w^\ell=0)$ has support $\{0\}$, and a function f on $(w^\ell=0)$ is determined by the numbers $f^{(j)}(0)$ for $0\le j<\ell$. For $\ell=2$, it is $\mathbb{D}$.

5

Tangent spaces and singular points

The tangent space

Let $X = (I = 0)$, with $I = (f_1, \ldots, f_N)$, and let P be a point of its support.

Definition: tangent space $T_P X$

$T_P X = \ker Df(P) = \{\, v \in \Bbbk^d : Df(P)\, v = 0 \,\}$, where $Df = (\partial f_i / \partial w_j)$ is the Jacobian matrix. Plane curve $(f = 0)$: $\nabla f(P) \cdot v = 0$.

Through $\mathbb{D}$: since $\varepsilon^2 = 0$ and $f_i(P) = 0$, we get $f_i(P + \varepsilon v) = \varepsilon\, Df_i(P)\, v$: the directions in which P can move, infinitesimally, and still satisfy the equations.
Loosely: maps from $\mathbb{D}$ to X are like a point on X with a tangent direction at it.

Definition: singular point; smooth point

P singular $\iff \dim T_P X >$ the dimension of the support near P; else smooth.
Plane curve $(f = 0)$: P singular $\iff \nabla f(P) = 0$.

Examples of singular points

	at	$\dim T_P$	
node $ab = 0$	0; elsewhere	2; 1	singular; smooth
cusp $b^2 = a^3$	0	2	singular
$(w = 0)$; $(w^2 = 0)$, on the line	0	0; 1	smooth; singular
double line $b^2 = 0$	every point	2	singular: $\nabla f = (0, 2b)$

Proposition: over $\mathbb{C}$, smooth implies reduced

If g^2 divides f and $g(P) = 0$, then $\nabla f(P) = 0$: $\nabla(g^2 h) = 2gh\,\nabla g + g^2\,\nabla h$. So over $\mathbb{C}$, a curve smooth at every point is *reduced*: f has no repeated factor.

A point can be singular in two ways: $(\sqrt{I} = 0)$ is already singular there (the node), or $(I = 0)$ has a larger tangent space than $(\sqrt{I} = 0)$ ($\mathbb{D}$, the double line).

Degenerate directions are tangent vectors

Definition: fibre of Φ through w_0

Write $f_w - f_{w_0} = \sum_{i=1}^{N} a_i(w)\, e_i$, with the e_i linearly independent functions of x and the a_i polynomials in w. The fibre is the scheme $(a_1 = \cdots = a_N = 0)$.

Its real points are the parameters with the same function as w_0. Its tangent space at w_0 is $\ker Da(w_0) = \{\, v : D_v\Phi(w_0) = 0 \,\}$, Lecture 1's degenerate directions ($\ker H(w_0)$ for the square loss). It is non-zero when the fibre (examples from Lecture 1):

has positive dimension	$f = a + b$: the fibres are lines
has a singular point	$f = ab$ at 0: dimension 1 jumps to 2
is non-reduced at w_0	$f = w^2$ or w^3 at 0: dimension 1 at a point

The kernel of the Hessian is this tangent space. It tells (w) from (w^2) but not (w^2) from (w^3), while the ideals tell all three apart.

Use the fibre of Φ, not the scheme $(K = 0)$

Warning

Since $K \geq 0$ and $K = 0$ on W_0, the gradient ∇K vanishes at every true parameter. So the scheme $(K = 0)$ is singular at every point of W_0.

- Even the regular model, with $K = \frac{1}{2}w^2$: $(K = 0)$ is $(w^2 = 0) = \mathbb{D}$, whose tangent space at 0 is the whole line.
- Its fibre, $(w = 0)$, is a reduced point, with tangent space 0.
- So degeneracy is read from the ideal $(a_1, \ldots, a_N)$ of the fibre, not from the ideal (K).

6

Curves and divisors

Curves and their support

Definition: curve in the plane

A scheme $(f = 0)$, with f non-constant, up to a constant $c \neq 0$.
Its *support* is the set $\mathbb{V}(f)$. *Reduced* if f has no repeated factor.

- $b^2 = 0$: supported on the line $\mathbb{V}(b)$; not reduced. Its *reduced curve* is the curve of the radical: $b = 0$, with the same support.
- Over $\mathbb{R}$, the curve $a^2 + b^2 = 0$ is supported at a point.

Divisors

Definition: divisor

A formal sum $\sum_i \ell_i D_i$, with the D_i irreducible reduced curves and $\ell_i \in \mathbb{Z}$. *Effective* if all $\ell_i \geq 0$.

Definition: prime divisor $\operatorname{div}(g)$; divisor $\operatorname{div} f$ of a function

For g irreducible: $\operatorname{div}(g)$ = the curve $g = 0$, counted once, with support $\mathbb{V}(g)$.
For $f = c\, g_1^{\ell_1} \cdots g_r^{\ell_r}$ with the g_i distinct and irreducible: $\operatorname{div} f = \sum_i \ell_i \operatorname{div}(g_i)$.

$\operatorname{div}(fg) = \operatorname{div} f + \operatorname{div} g$: $\operatorname{div}(a^2 b^2) = 2\operatorname{div}(a) + 2\operatorname{div}(b)$. Over $\mathbb{R}$, $\operatorname{div}\left((a^2+b^2)^2\right) = 2\operatorname{div}(a^2+b^2)$: one prime divisor, counted twice, supported at a point. Forgetting the ℓ_i is taking the radical. Curves in the plane = non-zero effective divisors.

Key point: what we shall resolve is a divisor, that is, a sum of irreducible reduced curves, each with a multiplicity.

7

Simple normal crossings

Simple normal crossings

Definition: simple normal crossings of $D = \sum_i \ell_i D_i$

Near each point, in suitable coordinates u:
$D = \sum_{i \le j} \ell_i \operatorname{div}(u_i) = \operatorname{div}\left(u_1^{\ell_1} \cdots u_j^{\ell_j}\right)$: a *monomial*, up to a factor $\neq 0$; at most d components in dimension d.

Plane: smooth components, different tangents, at most two through a point.



simple normal crossings



fails: tangency



fails: a triple point

Why monomials: integrals of a monomial over a box split variable by variable.

8

The projective line

The projective line

Definition: projective line $\mathbb{P}^1$

The lines through 0 in $\mathbb{A}^2$:

$$\mathbb{P}^1 = \big(\mathbb{A}^2 \setminus \{0\}\big)/\sim,$$

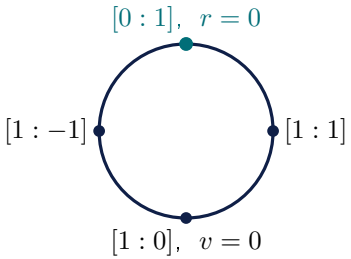
$(x, y) \sim (cx, cy) \quad \forall c \neq 0$. Its points are written $[x : y]$.

Definition: charts

$x \neq 0$: $v = y/x$. $\quad$ $y \neq 0$: $r = x/y$.

Glued by $r = 1/v$; each misses one point.

$\mathbb{P}^2$ and points at infinity: in the notes.



$\mathbb{RP}^1$ is a circle.

9

The blow-up of the plane

The projective line

Definition: projective line $\mathbb{P}^1$

The lines through 0 in $\mathbb{A}^2$:

$$\mathbb{P}^1 = \big(\mathbb{A}^2 \setminus \{0\}\big)/\sim,$$

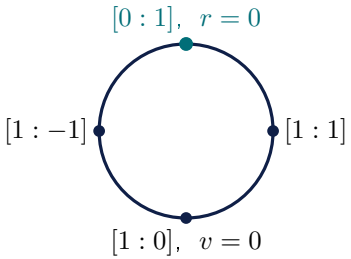
$(x, y) \sim (cx, cy) \quad \forall c \neq 0$. Its points are written $[x : y]$.

Definition: charts

$x \neq 0$: $v = y/x$. $\quad$ $y \neq 0$: $r = x/y$.

Glued by $r = 1/v$; each misses one point.

$\mathbb{P}^2$ and points at infinity: in the notes.



$\mathbb{RP}^1$ is a circle.

The blow-up of the plane

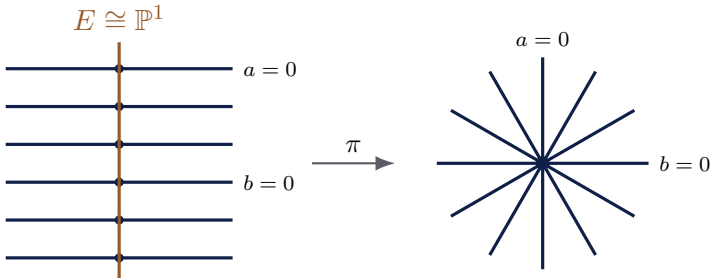
Definition: blow-up of the plane at 0

$$\mathrm{Bl}_0\,\mathbb{A}^2 = \left\{ \big((a,b),[x:y]\big) \in \mathbb{A}^2 \times \mathbb{P}^1 \;:\; ay = bx \right\} \xrightarrow{\pi} \mathbb{A}^2,$$

with π the projection to (a,b). $ay = bx$: the point (a,b) lies on the line $[x:y]$.

- Away from 0: exactly one $[x:y]$, namely $[a:b]$. So π is invertible there, with inverse $(a,b) \mapsto \big((a,b),[a:b]\big)$.
- Over 0: all of $\mathbb{P}^1$. The *exceptional divisor* E, supported on $\pi^{-1}(0) \cong \mathbb{P}^1$.
- A point of E is a direction through 0.

The blow-up separates directions



$\mathrm{Bl}_0\,\mathbb{A}^2$: the lines pulled apart

$\mathbb{A}^2$: all lines meet at 0

Lifted to the blow-up, each line through 0 meets E at its own point, so lines of different slopes no longer meet.

The two charts

Chart $x \neq 0$: coordinates (u, v)

$$\pi(u, v) = (u,\ uv)$$

$$E = \operatorname{div}(u), \qquad v = b/a \text{ off } E$$

$$D\pi = \begin{pmatrix} 1 & 0 \\ v & u \end{pmatrix}, \qquad \det D\pi = u$$

Chart $y \neq 0$: coordinates (r, s)

$$\pi(r, s) = (rs,\ s)$$

$$E = \operatorname{div}(s), \qquad r = a/b \text{ off } E$$

$$D\pi = \begin{pmatrix} s & r \\ 0 & 1 \end{pmatrix}, \qquad \det D\pi = s$$

In each chart, $\det D\pi$ is the equation of E: $\operatorname{div}(\det D\pi) = E$.

The gluing

On the overlap, where $v \neq 0$ and $r \neq 0$:

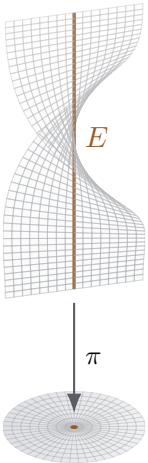
$$r = \tfrac{1}{v}, \quad s = uv, \qquad \text{with inverse} \qquad u = rs, \quad v = \tfrac{1}{r}.$$

The two formulas for π agree: $(rs,\ s) = \left(\tfrac{1}{v} \cdot uv,\ uv\right) = (u,\ uv)$, and

$$\frac{\partial(r,s)}{\partial(u,v)} = \begin{pmatrix} 0 & -1/v^2 \\ v & u \end{pmatrix}, \qquad \det = \frac{1}{v} \neq 0.$$

On E the gluing is $v \mapsto 1/v$: the two charts of $\mathbb{P}^1$.

Over the reals: a Möbius band



- $E \cong \mathbb{P}^1(\mathbb{R})$ is the vertical axis; height along it is a direction through 0.
- Through each point of E runs the lift of the line with that direction. It turns through a half-turn from bottom to top.
- π is vertical projection to the plane below.
- The two ends of E are the same point of $\mathbb{P}^1(\mathbb{R})$. Glued, a neighbourhood of E is an open Möbius band.

Divisors

Definition: divisor

A formal sum $\sum_i \ell_i D_i$, with the D_i irreducible reduced curves and $\ell_i \in \mathbb{Z}$. *Effective* if all $\ell_i \geq 0$.

Definition: prime divisor $\operatorname{div}(g)$; divisor $\operatorname{div} f$ of a function

For g irreducible: $\operatorname{div}(g)$ = the curve $g = 0$, counted once, with support $\mathbb{V}(g)$.
For $f = c\, g_1^{\ell_1} \cdots g_r^{\ell_r}$ with the g_i distinct and irreducible: $\operatorname{div} f = \sum_i \ell_i \operatorname{div}(g_i)$.

$\operatorname{div}(fg) = \operatorname{div} f + \operatorname{div} g$: $\operatorname{div}(a^2 b^2) = 2\operatorname{div}(a) + 2\operatorname{div}(b)$. Over $\mathbb{R}$, $\operatorname{div}\left((a^2+b^2)^2\right) = 2\operatorname{div}(a^2+b^2)$: one prime divisor, counted twice, supported at a point. Forgetting the ℓ_i is taking the radical. Curves in the plane = non-zero effective divisors.

Key point: what we shall resolve is a divisor, that is, a sum of irreducible reduced curves, each with a multiplicity.

Strict and total transforms

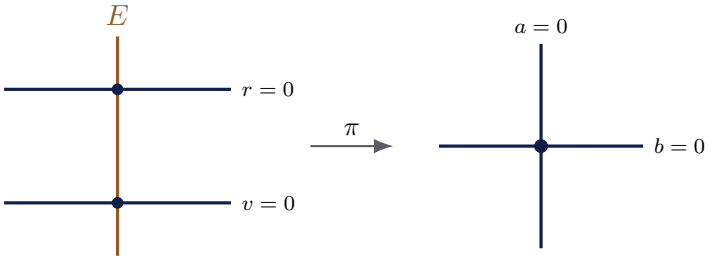
$D = \operatorname{div}(f)$, the curve $f = 0$, through 0. The multiplicity $m = \operatorname{mult}_0 f$ is the lowest degree of a monomial of f.

Definition: strict transform $\widetilde{D}$, in the chart (u, v)

1. Substitute $(a, b) = (u, uv)$.
2. Factor: $f(u, uv) = u^m \tilde{f}(u, v)$, with u not dividing $\tilde{f}$.
3. Keep the rest: the *strict transform* is $\widetilde{D} = \operatorname{div}(\tilde{f})$, supported on $\{\tilde{f} = 0\}$.

$\widetilde{D}$ meets E where $\tilde{f}(0, v) = 0$: the tangent directions of D at 0. As divisors, the *total transform* is $\operatorname{div}(f \circ \pi) = m\,E + \widetilde{D}$. Likewise in the chart (r, s), with $f(rs, s) = s^m \tilde{f}(r, s)$.

The node: one blow-up separates its branches



$\widetilde{D}$ meets E at two points

$\mathbb{V}(ab)$: the branches cross

Chart (u, v): $ab \circ \pi = u^2 v$.

Chart (r, s): $ab \circ \pi = s^2 r$.

$\operatorname{div}(ab \circ \pi) = 2E + \widetilde{D}$: the two branches meet E at two different points.

$\operatorname{div}(ab)$ already had simple normal crossings.

10

An example: the cusp

Simple normal crossings

Definition: simple normal crossings of $D = \sum_i \ell_i D_i$

Near each point, in suitable coordinates u:
$D = \sum_{i \le j} \ell_i \operatorname{div}(u_i) = \operatorname{div}\left(u_1^{\ell_1} \cdots u_j^{\ell_j}\right)$: a *monomial*, up to a factor $\neq 0$; at most d components in dimension d.

Plane: smooth components, different tangents, at most two through a point.



simple normal crossings



fails: tangency



fails: a triple point

Why monomials: integrals of a monomial over a box split variable by variable.

The two charts

Chart $x \neq 0$: coordinates (u, v)

$$\pi(u, v) = (u,\ uv)$$

$$E = \operatorname{div}(u), \qquad v = b/a \text{ off } E$$

$$D\pi = \begin{pmatrix} 1 & 0 \\ v & u \end{pmatrix}, \qquad \det D\pi = u$$

Chart $y \neq 0$: coordinates (r, s)

$$\pi(r, s) = (rs,\ s)$$

$$E = \operatorname{div}(s), \qquad r = a/b \text{ off } E$$

$$D\pi = \begin{pmatrix} s & r \\ 0 & 1 \end{pmatrix}, \qquad \det D\pi = s$$

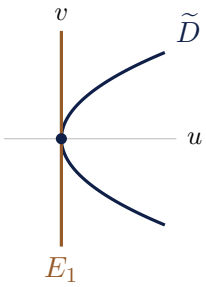
In each chart, $\det D\pi$ is the equation of E: $\operatorname{div}(\det D\pi) = E$.

The cusp: first blow-up

$f = b^2 - a^3$, $m = 2$. In the chart (u, v):

$$f \circ \pi_1 = u^2 v^2 - u^3 = u^2 (v^2 - u),$$

$$\operatorname{div}(f \circ \pi_1) = 2E_1 + \widetilde{D}.$$



- $E_1 = \operatorname{div}(u)$; $\widetilde{D} = \operatorname{div}(u - v^2)$ is smooth, the parabola $u = v^2$.
- But $\tilde{f}(0, v) = v^2$ has a double root: $\widetilde{D}$ is tangent to E_1.

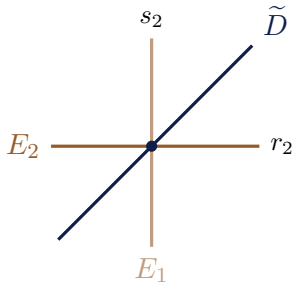
So this is not yet normal crossings.

The cusp: second blow-up

Blow up the origin of (u, v), in the chart $u = r_2 s_2$, $v = s_2$. Then $v^2 - u = s_2(s_2 - r_2)$, so

$$f \circ \pi_1 \circ \pi_2 = r_2^2 \, s_2^3 \, (s_2 - r_2).$$

$E_2 = \mathrm{div}(s_2)$, $E_1 = \mathrm{div}(r_2)$, $\widetilde{D} = \mathrm{div}(s_2 - r_2)$: three smooth curves through one point, with three different tangents. A triple point.



The cusp: third blow-up

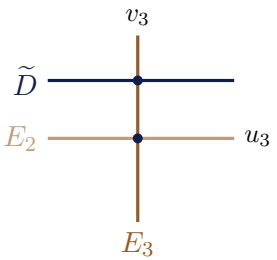
Blow up the origin of (r_2, s_2), in the chart $r_2 = u_3$, $s_2 = u_3 v_3$. With $g = \pi_1 \circ \pi_2 \circ \pi_3$, we get $f \circ g = u_3^6 \, v_3^3 \, (v_3 - 1)$: $E_3 = \operatorname{div}(u_3)$, $E_2 = \operatorname{div}(v_3)$, $\widetilde{D} = \operatorname{div}(v_3 - 1)$; in the other chart, E_1 crosses E_3.

Simple normal crossings:

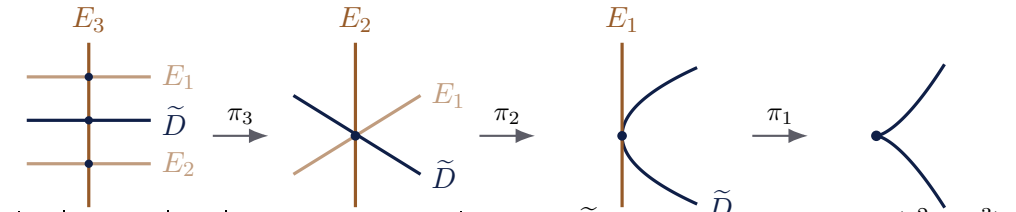
$$\operatorname{div}(f \circ g) = 2E_1 + 3E_2 + 6E_3 + \widetilde{D}.$$

Jacobians multiply (the chain rule), so their divisors add: $\det Dg = u \cdot s_2 \cdot u_3 = u_3^4 \, v_3^2$, and

$$\operatorname{div}(\det Dg) = E_1 + 2E_2 + 4E_3.$$



The cusp: three blow-ups to simple normal crossings



$$\operatorname{div}(f \circ g) = 2E_1 + 3E_2 + 6E_3 + \widetilde{D}, \qquad \operatorname{div}(\det Dg) = E_1 + 2E_2 + 4E_3.$$

One blow-up makes the strict transform smooth; simple normal crossings take three. We resolve the curve as a divisor in the plane, not as a curve on its own: the two later blow-ups remove a tangency, then a triple point.

11

Resolution of singularities

The map Φ and the loss K

The map $\Phi : \mathcal{W} \to \mathcal{F}$, $w \mapsto f_w$

From parameters to functions of x. A map has no equations; its fibre through w_0 is the scheme $(a_1 = \cdots = a_N = 0)$, where $f_w - f_{w_0} = \sum_i a_i(w)\, e_i$.

The loss $K(w) = L(w) - L(w_0) \geq 0$

One function of w; the true parameters are $W_0 = \{K = 0\}$. For the square loss, with $w_0 \in W_0$ and a positive definite Gram matrix $M_{ij} = \mathbb{E}\big[e_i(x)\, e_j(x)\big]$,

$$K(w) = \tfrac{1}{2} \sum_{i,j} a_i(w)\, M_{ij}\, a_j(w),$$

so K vanishes exactly where every a_i does.

This section resolves the divisor of the single function K.

Hironaka's theorem for polynomials

For a polynomial f, finitely many blow-ups like today's give a map g for which $\operatorname{div}(f \circ g)$ has simple normal crossings (Hironaka 1964). In dimension 3 and up, a blow-up may also be along a smooth curve or surface lying over the zero set of f.

- Near each point, in suitable coordinates, $f \circ g$ is a monomial times a factor not zero there: near $(0,1)$, the cusp's $u_3^6 v_3^3(v_3 - 1)$ is $u_3^6(v_3 - 1)$ times $v_3^3 \neq 0$.
- So is $\det Dg$: each $\det D\pi$ is a power of an equation of E.

Key point: for SLT's K we resolve its divisor, not its zero set

The divisor keeps the multiplicities. Since $K \geq 0$, every point of its zero set is a singular point of the scheme $(K = 0)$. Over $\mathbb{R}$, $K = (a^2 + b^2)^2$ vanishes only at 0; one blow-up gives $u^4(1 + v^2)^2$, which vanishes on all of $\pi^{-1}(0)$, the support of E: the point has become a curve.

Watanabe's form of the theorem

Theorem (Watanabe 2009)

Near each true parameter w_0 there is a map g, from a new space onto a neighbourhood of w_0, invertible away from the zero set of K. Around each point of the new space there are coordinates u in which

$$K\big(g(u)\big) = u_1^{2k_1} \cdots u_d^{2k_d}, \qquad \big|\det Dg(u)\big| = b(u)\,\big|u_1^{h_1} \cdots u_d^{h_d}\big|,$$

with $b > 0$, and exponents that depend on the point.

For polynomial K, Hironaka's blow-ups give such a g.

Example: the cusp's $K = (b^2 - a^3)^2$

Near $(u_3, v_3) = (0, 1)$, with $t = v_3 - 1$: $K \circ g = u_3^{12}\, t^2 \cdot v_3^6$ and $|\det Dg| = v_3^2\, u_3^4$, where $v_3 \neq 0$. One more change of coordinates absorbs the factor v_3^6.

Real analytic K and computing g

Definition: real analytic function K

Near each point, K is a convergent power series. Example: the K of a network with $\tanh$ activations and bounded inputs.

For real analytic K the theorem holds as stated (Watanabe 2009), and g can still be built from blow-ups (Bierstone and Milman 1997).

Computing g

The proofs of the theorem give algorithms (Villamayor 1989; Bierstone and Milman 1997), but resolving a given K is hard in practice. Every example today is polynomial.

The precise statement is in the notes.

From local to global

- Each point of a compact parameter space has a neighbourhood with its own resolution, by finitely many blow-ups.
- Finitely many such neighbourhoods cover the space, and Lecture 3 splits the integral into one piece for each neighbourhood.
- Neighbourhoods, not singular points: every zero of K is a singular point of the scheme $(K = 0)$, and the true parameters can form a curve or a surface.

When polynomials are enough

SLT's K is rarely a polynomial; resolution depends only on how K vanishes.

Definition: a polynomial Q equivalent to K near w_0

$c_1 Q \leq K \leq c_2 Q$ near w_0, for constants $c_1, c_2 > 0$.

Then the blow-ups that resolve Q also resolve K: $K \circ g$ is the same monomial, times a positive factor. This happens often:

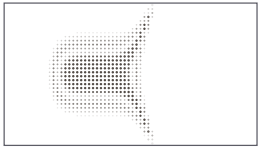
- models polynomial in w (ab, w^3, deep linear networks): K is a polynomial;
- polynomials times factors not zero at w_0: $\tanh(a)\tanh(b) = ab\times$ (such a factor);
- an isolated true parameter: Q a Taylor polynomial of K, of high enough degree.

Not always: no polynomial $Q \neq 0$ vanishes along $b = \sin a$, so none is equivalent to $(b - \sin a)^2$. Then the analytic theorem applies directly, and blow-ups still work.

12

Summary

What last week's pictures show



- Last week's title: the loss $K = (b^2 - a^5)^2$ from above, bands of equal height lit in turn. At the bottom (left): its zero set, the curve $b^2 = a^5$, singular at 0.
- The ornaments resolve that curve: one blow-up per arrow, one chart per picture, as for our cusp. They resolve K too.



$b^2 = a^5$ | a cusp | tangent to E_1 | a triple point | simple normal crossings

Pictures: Matthew Farrugia-Roberts. Ornaments after Franklin Vera Pacheco, *Functorial Resolution of Singularities* (Wikimedia Commons), CC BY-SA 3.0.

Summary

- Vanishing locus. $\mathbb{V}(I)$ is a set; the scheme $(I = 0)$ keeps the ideal I, and with it what the set loses, such as the order of vanishing.
- Simple normal crossings. Smooth components meeting like coordinate hyperplanes: a monomial local equation.
- Resolution of singularities. Finitely many blow-ups give a map g and coordinates in which $K \circ g$ is a monomial; $\det Dg$ is then a monomial times a factor that is not zero.

Reserve

Reserve: the Zariski topology

- The closed sets are the vanishing loci $\mathbb{V}(I)$: finite unions and arbitrary intersections of them are again vanishing loci.
- In $\mathbb{A}^1$ the closed sets are the finite sets and $\mathbb{A}^1$. Any two non-empty open sets meet: not Hausdorff.
- It is bookkeeping, not analysis: it makes irreducibility a topological notion.

Reserve: tangent cones and the strict transform

Split f into homogeneous pieces, $f = f_{[m]} + f_{[m+1]} + \cdots$ with $f_{[m]} \neq 0$, so $m = \operatorname{mult}_0 f$; the lowest piece $f_{[m]}$ is the *tangent cone*. In the chart $\pi(u, v) = (u, uv)$,

$$f(u, uv) = u^m \tilde{f}(u, v), \qquad \tilde{f}(u, v) = f_{[m]}(1, v) + u\, f_{[m+1]}(1, v) + \cdots$$

So $\widetilde{D}$ meets $E = \operatorname{div}(u)$ where $f_{[m]}(1, v) = 0$: the tangent directions of $f = 0$ at 0, with multiplicity.

node ab	$f_{[2]}(1, v) = v$	a simple root, and one in (r, s)
nodal cubic $b^2 = a^2(a + 1)$	$f_{[2]}(1, v) = v^2 - 1$	two simple roots
cusp $b^2 = a^3$	$f_{[2]}(1, v) = v^2$	one double root

Reserve: divisors determine principal ideals only

The scheme $(a^2 = ab = 0)$, with ideal $(a^2, ab) = a \cdot (a, b)$, agrees with the line $(a = 0)$ away from the origin, with extra structure at the origin.

- Its support is the line $\mathbb{V}(a)$, and the divisors of its elements are all at least $\mathrm{div}(a)$: neither detects the extra structure.
- The tangent space detects it: dimension 2 at the origin, 1 elsewhere on the line. The scheme is singular at the origin only.
- This is why the general theorem is stated for ideals (principalisation): an ideal with several generators carries structure that no divisor records.
- For one function, as K, the ideal (K) and the divisor $\mathrm{div}(K)$ carry the same information.

Reserve: properties of ord_D

- Near a smooth point of an irreducible reduced curve D, with local equation $t = 0$, every function is uniquely t^ℓ times a factor that is not zero on D there; then $\mathrm{ord}_D(f) = \ell$.
- So $\mathrm{ord}_D(fg) = \mathrm{ord}_D(f) + \mathrm{ord}_D(g)$, and $\mathrm{ord}_D(f+g) \geq \min\big(\mathrm{ord}_D(f), \mathrm{ord}_D(g)\big)$. A map with these two properties is called a *valuation*.
- This makes sense for analytic f, where "g^ℓ divides f" no longer parses: it is what makes $\mathrm{div}(K)$ work for analytic K.

Reserve: three versions of Hironaka's theorem

Over $\mathbb{R}$ or $\mathbb{C}$, by finitely many blow-ups along smooth centres (Hironaka 1964):

1. A smooth model. For a *reduced* scheme X (its ring has no non-zero nilpotents): a map $Y \to X$ from a smooth Y that is *proper* (preimages of compact sets are compact) and *birational* (an isomorphism away from lower-dimensional sets).
2. Embedded. A divisor, together with the exceptional divisors, brought to simple normal crossings.
3. Principalisation. A non-zero ideal made *principal and monomial*: near each point, generated by one monomial in suitable coordinates.

Reserve: the precise statement of Watanabe's theorem

Theorem (after Watanabe 2009)

Let $K \geq 0$ be real analytic near $w_0 \in \mathcal{W}$, with $K(w_0) = 0$, and not identically zero near w_0. There are an open neighbourhood U of w_0, a real analytic manifold Y of dimension d, and a proper real analytic map $g : Y \to U$, which is an analytic isomorphism from $Y \setminus g^{-1}(W_0)$ onto $U \setminus W_0$, such that around each point P of $g^{-1}(W_0)$ there are local coordinates $u = (u_1, \ldots, u_d)$, centred at P, in which

$$K\big(g(u)\big) = u^{2k} = u_1^{2k_1} \cdots u_d^{2k_d}, \qquad \big|\det Dg(u)\big| = b(u)\,\big|u^h\big|,$$

where k and h are multi-indices in $\mathbb{N}^d$, depending on P, and $b > 0$ is real analytic.

Here $W_0 = \{K = 0\}$. The theorem is local: one neighbourhood U at a time.

Reserve: the precise statement on blow-ups

Blow-ups

For polynomial K, g can be taken to be a composite of finitely many blow-ups along smooth centres lying over the zero set of K: this is Hironaka's theorem (Hironaka 1964; Watanabe 2009). The same holds for real analytic K over a relatively compact neighbourhood of w_0 (Bierstone and Milman 1997, for real-analytic manifolds, non-compact domains included).

An open set is *relatively compact* if its closure is compact.

Reserve: when a polynomial is enough

Call real analytic $K_1, K_2 \geq 0$ *equivalent* near w_0 if $c_1 K_2 \leq K_1 \leq c_2 K_2$ there, for constants $c_1, c_2 > 0$ (Watanabe 2009). Then every resolution of K_2 resolves K_1, with the same exponents. An analytic K is equivalent to a polynomial, for instance:

- when the coefficients of $\Phi(w) - \Phi(w_0)$, in a basis of functions of x, are polynomials;
- for a polynomial times a factor not zero near w_0: $\tanh(a)\tanh(b)$ is ab times such a factor;
- near an isolated zero (Łojasiewicz's inequality).

Not always: no polynomial is equivalent to $(b - \sin a)^2$; then use the analytic theorem. Details: in the notes.

Many thanks for your attention!

Any questions?