О работах В. И. Арнольда по теории особенностей

С. М. Гусейн-Заде

В этом году исполняется 80 лет со дня рождения одного из крупнейших математиков XX века Владимира Игоревича Арнольда. Ему принадлежат работы в разных областях математики, и дать сколько-нибудь полный обзор его математической деятельности выше моих возможностей. Моя научная деятельность, примыкающая к работам Владимира Игоревича, концентрировалась вокруг теории особенностей. Данная статья — это эссе о некоторых моментах его работы в этой области.

Теория особенностей как отдельная дисциплина возникла в конце 60-х — начале 70-х годов прошлого века. Конечно, и до этого был получен ряд результатов, которые естественно составляют её часть. Сюда можно отнести, в частности, лемму Морса, классификацию Уитни общих отображений плоскости в плоскость... Однако систематическое изучение объектов теории особенностей началось с работ Рене Тома (René Thom) и получило наиболее значительное развитие в работах Арнольда. При этом, как мне кажется, идеология Арнольда существенно отличалась от идеологии его предшественников, в том числе и Тома.

Многие из них подходили к задачам теории особенностей как к чисто классификационным. Так, Том исходил из задачи описания перестроек фазовых портретов градиентных динамических систем в общих семействах систем, зависящих от не более чем четырёх параметров. (Выбор такого количества параметров связывался с тем, что рассматриваемая система могла зависеть от точки пространства и от времени.) Поэтому он изучал семейства функций (потенциалов) общего положения, зависящие от указанного количества параметров. С точки зрения строения фазовых портретов соответствующих динамических систем (по крайней мере с «локальной» точки зрения) наибольший интерес представляют окрестности критических точек потенциалов. В окрестности некритической точки потенциала градиентная динамическая система устроена очень просто. Именно, она может быть «выпрямлена». Это означает, что в некоторых локальных координатах $x_1, \ldots, x_n$ она имеет вид $dx_1/dt = 1$, $dx_i/dt = 0$ при $2 \leqslant i \leqslant n$. В этих координатах интегральные кривые системы образуют семейство параллельных прямых. Эта картина устойчива, т. е. не меняется при небольших изменениях потенциала (или при небольших изменениях параметров, от которых он зависит).

У индивидуального потенциала, т. е. для «семейства, зависящего от нуля параметров», все критические точки невырожденны. Это означает, что, в некоторых локальных координатах $x_1, \ldots, x_n$, в которых критическая точка имеет нулевые координаты, потенциал имеет вид $c + \sum_{i=1}^{n} \pm x_i^2$. (Невырожденные критические точки, в которых в этой формуле присутствуют только плюсы, являются положениями (устойчивого) равновесия.) Интегральные кривые соответствующей градиентной динамической системы могут быть легко описаны, и их картина качественно не меняется при небольших изменениях потенциала.

Ситуация несколько другая для семейств потенциалов, зависящих от одного параметра. Для общего значения параметра (т. е. для почти всех его значений) потенциал опять-таки имеет только невырожденные критические точки. При этих исключительных (бифуркационных) значениях параметра критические точки устроены по-другому. Можно показать, что для однопараметрического семейства общего положения все критические точки потенциала, соответствующего бифуркационному значению параметра, невырожденны, за исключением одной, а эта одна в некоторых локальных координатах имеет вид $c + x_1^3 + \sum_{i=2}^{n} \pm x_i^2$. При этом с одной стороны от исключительного значения параметра потенциал имеет около этой критической точки две невырожденные критические точки, которые сливаются при исключительном значении параметра и исчезают с другой стороны от исключительного значения (можно в определённом смысле сказать, что они уходят в комплексную область). Если одна из указанных двух критических точек была минимумом потенциала (и значит, устойчивым положением равновесия) и система находилась в нём, то при прохождении параметра через бифуркационное значение этот минимум исчезает и система должна перейти в другое устойчивое состояние с меньшей потенциальной энергией. Если система достаточно быстрая, такой переход осуществляется скачком. В соответствии с законом сохранения энергии в этом случае выделяется энергия (равная разности значений потенциала в двух положениях равновесия: до перехода и после). Можно сказать, что система испытывает быструю перестройку: происходит катастрофа.

Иногда на своих лекциях я демонстрирую этот эффект с помощью некоторой игрушки, которую я получил на какой-то выставке на ВДНХ. Игрушка представляет собой немного выпуклый металлический диск с наклеенным на него пластиковым кружочком, на котором написано (довольно мелкими буквами) название фирмы. На выставке работники фирмы что-то делали с таким диском, после чего клали его на стойку. Для того чтобы увидеть, что написано на диске, посетители выставки наклонялись над ним. В этот момент диск неожиданно подпрыгивал, пугая посетителей. Описанный эффект объясняется следующим образом. Центральную часть диска можно прогибать, нажимая на неё. Потенциальная энергия диска зависит от величины прогиба. При этом она зависит ещё от одного параметра: от температуры диска. При комнатной температуре потенциальная энергия как функция от величины прогиба имеет один минимум и система (диск) находится в соответствующем положении равновесия. При несколько повышенной температуре (ближе к температуре тела человека) имеются два локальных минимума потенциальной энергии как функции от величины прогиба (и ещё один локальный максимум). Перед тем как положить диск на стойку, работник фирмы нагревал его пальцами, в результате чего появлялось второе устойчивое положение равновесия. После этого прикладывая некоторое усилие, надавливая на середину диска и тем самым совершая работу, он переводил его во вновь появившееся положение равновесия и клал на стойку. Там диск охлаждался до комнатной температуры, новый локальный минимум потенциальной энергии сливался с локальным максимумом и исчезал. Система (диск) скачком переходила в единственное положение равновесия, существующее при комнатной температуре. При этом выделялась энергия, ранее затраченная для перевода системы в новое положение равновесия, и диск подпрыгивал.

Том изучил ситуацию для семейств потенциалов общего положения, зависящих от не более чем четырёх параметров, и обнаружил, что имеются ещё шесть возможных локальных ситуаций (с точностью до некоторого естественного отношения эквивалентности: см. ниже). Тем самым он утверждал, что в общих семействах градиентных динамических систем, зависящих от четырёх или менее параметров, имеется семь возможных типов перестроек («катастроф»). Теория, начинающаяся с этого утверждения, стала называться теорией катастроф. Она получила довольно большую популярность. В её популяризации одну из ключевых ролей сыграли общедоступные статьи и книги К. Зимана (Ch. Zeeman). Утверждалось, что теория катастроф — это универсальная теория, описывающая резкие (не плавные) изменения систем. При этом делались попытки применять её повсюду. Такого рода применение можно найти в брошюре Арнольда «Теория катастроф». Сам термин «теория катастроф» Владимиру Игоревичу не очень нравился. Тот факт, что он так назвал брошюру, видимо, объясняется популярностью названия в обществе.

Надо заметить, что как математическое утверждение описанный результат Тома неверен. Дело в том, что градиентная динамическая система определяется не только своим потенциалом (как функцией от фазовых переменных), но и римановой метрикой на фазовом пространстве. Поэтому, строго говоря, речь должна идти об изучении (классификации) не просто критических точек функций в семействах, а пар, состоящих из функции и метрики. Такая классификация существенно сложнее и имеет гораздо больше семи классов. Я не знаю, какова ситуация вокруг неё в настоящее время, но некоторое время тому назад было известно, что существует как минимум 27 классов эквивалентности, и не было известно, конечно ли их общее число.

Свои исследования по теории особенностей Арнольд тоже начинал с задачи, в которой естественной была бы классификация критических точек функций, встречающихся в семействах общего положения, зависящих от малого числа параметров. Речь шла о следующей задаче. Рассматривается «быстро осциллирующий интеграл» следующего вида:

$$\int\limits_{\mathbb{R}^n} \varphi(x_1, \ldots, x_n) e^{f(x_1,\ldots,x_n)/h} dx_1 \ldots dx_n,$$

где $\varphi(\overline{x})$ — функция с компактным носителем (нужная только для того, чтобы интеграл имел смысл; от неё ничего существенного не зависит), h — малая величина (в приложениях часто — «постоянная Планка»). Можно показать, что если на области интегрирования — носителе функции $\varphi(\overline{x})$ — фаза $f(\overline{x})$ не имеет критических точек, то при h, стремящемся к нулю, значение интеграла стремится к нулю быстрее любой степени h. Таким образом, нетривиальные степенные асимптотики связаны с критическими значениями функции f. Это утверждение называется принципом стационарной фазы. Итак, задача формально сводится к классификации критических точек, которые могут встречаться в общих семействах, зависящих от фиксированного числа параметров, и вычислению соответствующих показателей для них.

До коразмерности четыре классификация была проведена Томом. (Кстати, отмечу, что в отличие от задачи, рассматриваемой Томом, в рассматриваемой Арнольдом задаче действительно нужно классифицировать только критические точки функций, а не критические точки вместе с какими-то дополнительными структурами, например с римановой метрикой. В этом смысле математический результат Арнольда соответствует естественно-научной задаче, бывшей для него мотивировкой.) Более-менее в то же время, когда задачей классификации критических точек функций заинтересовался Арнольд, эта классификация была продвинута немного дальше.

Поэтому изначально решаемая Арнольдом задача сводилась к вычислению показателей асимптотики интеграла для уже расклассифицированных особенностей, и на этом тема была бы закрыта. Подозреваю, что такая работа привлекла бы внимание нескольких специалистов по методу стационарной фазы и не была бы известна никому, кроме них. Однако Владимир Игоревич решил сделать что-то совершенно другое.

Прежде чем описать то, что он сделал, отмечу одну особенность подхода Арнольда к математическим задачам, которая представляется мне мистической, объяснения которой у меня нет, и я никогда не мог её воспроизвести. Моё описание отражает только то, что видел и понимал я, а в действительности Владимир Игоревич видел и понимал существенно больше (в чём я не сомневаюсь), и поэтому, возможно, ничего мистического в этом не было. Однако я воспринимал и воспринимаю некоторые его подходы как совершенно необъяснимые. Именно, несколько раз на моей памяти он задавал вопрос или формулировал гипотезу, для которых, как казалось мне, да и другим участникам семинара тоже, никакого основания не было. Поэтому вопрос и/или гипотеза выглядели абсолютно ничем не мотивированными, и было неясно, откуда они взялись и что в них толку. Однако позже неожиданно оказывалось, что эти вопросы и гипотезы приводили к глубоким и интересным результатам и проблемам.

В обсуждаемой задаче также проявилась эффективность такого подхода Владимира Игоревича. Он решил, что надо классифицировать критические точки функций не по коразмерности, а по другой характеристике. Классификация Тома давала конечный список особенностей (классов эквивалентности), поскольку она заходила не слишком далеко. Начиная с некоторого места классификация перестаёт быть дискретной: классы эквивалентности особенностей могут зависеть от непрерывных параметров — модулей. Например, в семействах функций от двух переменных, зависящих от семи параметров, могут встречаться критические точки, которые в некоторых локальных координатах x, y имеют вид $xy(x+y)(x-\lambda y)$ с каким-нибудь (определённым!) значением параметра λ. При этом указанные критические точки с различными значениями параметра λ, вообще говоря, не эквивалентны, т. е. не могут быть получены друг из друга заменой координат. Это следует из того известного факта, что четыре прямые на плоскости, пересекающиеся в одной точке, имеют инвариант (модуль) — двойное отношение их коэффициентов наклона. (Эти четыре прямые — компоненты множества нулей функции.) Для функций трёх переменных подобный эффект проявляется ещё раньше: в семействах, зависящих от шести параметров.

Владимир Игоревич решил классифицировать (с точностью до замен координат) критические точки функций по количеству параметров (модулей), требуемых для их описания (более точно — для описания всех близких к ним функций). При таком подходе начинать надо с функций, описание которых модулей не требует. Такие функции Арнольд назвал простыми или 0-модальными. Сама по себе классификация таких функций не слишком сложна. (Конечно, это — апостериори. Многие математические утверждения, будучи сформулированы и доказаны, вскоре выглядят несложными. Проблема состоит в том, чтобы найти правильную формулировку и первый раз доказать.) Ответ зависит (не сильно) от того, какие функции (и, соответственно, какие замены переменных) рассматриваются: комплексно-аналитические (голоморфные) или вещественные (бесконечно) дифференцируемые. Сформулируем его в чуть более простом случае — для голоморфных функций. (Для вещественных функций некоторые из указанных классов эквивалентности расщепляются на подклассы, различающиеся знаками перед мономами. Например, голоморфная функция $x^2 + y^2$ в вещественном случае представлена функциями $x^2 + y^2$, $x^2 - y^2$ и $-x^2 - y^2$.) Для двух переменных такие функции (классы эквивалентности относительно замен координат) образуют две бесконечные серии $x^{k+1} + y^2$ $(k \geqslant 2)$ и $x^2y + y^{k-1}$ $(k \geqslant 4)$, помимо которых имеются три исключительных особенности: $x^3 + y^4$, $x^3 + xy^3$ и $x^3 + y^5$. При фиксированном количестве переменных n (не менее двух) количество классов эквивалентности такое же и они получаются из указанных функций двух переменных добавлением суммы квадратов остальных $(n-2)$ переменных. Вообще, классификации критических точек функций разного количества переменных тесно связаны между собой. Более точно, добавление к функции квадратов дополнительных переменных определяет отображение, а именно вложение, множества классов эквивалентности функций меньшего числа переменных в множество классов эквивалентности функций большего числа переменных. Это — следствие так называемой леммы Морса с параметрами.

После того как этот список был получен, оказалось, что он встречался уже не раз в других ипостасях. Например, для функций трёх переменных из этого списка множество нулей известно как множество особенностей дю Валя, т. е. особенностей поверхностей, получаемых из комплексной плоскости факторизацией по дискретным подгруппам группы $SU(2)$ унитарных матриц с определителем 1 (соответствующих правильным многогранникам). Они же дают список так называемых рациональных двойных особых точек поверхностей. Было известно соответствие этих особенностей простым алгебрам Ли, имеющим корни одинаковой длины. Это алгебры традиционно обозначаются A_k, D_k, E_6, E_7 и E_8. Те же имена были присвоены и перечисленным выше критическим точкам функций.

Позже были найдены ещё несколько классификаций, дающих тот же список. Комплексно-аналитической функции f, определённой в окрестности начала координат в $\mathbb{C}^n$, соответствует некоторое комплексное многообразие с краем — её локальное множество уровня. Именно, для достаточно маленького $\delta > 0$ существует такое $\varepsilon_0 > 0$, что для ненулевого комплексного числа ε, по модулю меньшего ε_0, множество

$$V_f = \{\bar{z} \in \mathbb{C}^n \colon \|\bar{z}\| \leqslant \delta, f(\bar{z}) = \varepsilon\}$$

является гладким многообразием с краем. Многообразие V_f, конечно, зависит от δ и от ε, но для достаточно маленьких δ и ε_0 все эти многообразия диффеоморфны друг другу, т. е. V_f как абстрактное гладкое многообразие размерности $2(n-1)$ с краем корректно определено. (Поэтому в обозначении обычно участвует только сама функция f.) Оно называется слоем Милнора функции f. (Откуда берётся термин «слой» — другая история.) Теорема того же Дж. Милнора утверждает, что если критическая точка f в начале координат изолирована (а только такие и встречаются в случае конечной коразмерности), слой Милнора V_f гомотопически эквивалентен букету сфер размерности $n-1$, т. е. может быть стянут на такой букет. Поэтому нетривиальные гомологии у этого многообразия имеются только в размерности $n-1$. Как для всякого (ориентированного) многообразия чётной размерности, на группе гомологий средней размерности имеется билинейная форма: форма пересечения. Она симметрична при нечётных n и кососимметрична при n чётных. Поскольку V_f — многообразие с краем, эта форма не обязана быть невырожденной. Так вот, если мы зададимся вопросом, для каких функций трёх переменных эта форма (а в этом случае она симметрична) отрицательно определена, мы получим тот же список A_k, D_k, E_6, E_7, E_8. Имеются и другие критерии, определяющие этот список. Я помню, что читал статью, в которой приводилось около двадцати классификаций, приводящих к тому же списку.

Удивительно, что до сих пор непонятны априорные причины, объясняющие, почему классификация простых функций совпадает с классификацией простых алгебр Ли, имеющих корни одинаковой длины, а также с классификацией функций трёх переменных, для которых указанная выше билинейная форма отрицательно определена. Также непонятна связь с некоторыми другими классификациями, дающими тот же ответ. Их совпадение получается просто независимой классификацией тех и других с последующим сравнением полученных списков. Это совпадение классификаций получало название ADE-загадки (или даже ADE-мистики) и ждёт своего объяснения.

Классификация простых ростков функций связана с классификацией не всех простых алгебр Ли, а только тех, которые имеют корни одинаковой длины. Владимир Игоревич нашёл классификации, приводящие к другим простым алгебрам Ли. Так, оказалось, что простым алгебрам Ли B_k, C_k и F_4 с корнями двух длин (1 и $\sqrt{2}$) отвечает классификация критических точек функций на многообразиях с краем.

Следующим шагом была классификация унимодальных критических точек функций, т. е. критических точек, для описания которых требуется один параметр (модуль). Примером таких функций являются указанные выше функции $xy(x+y)(x-\lambda y)$, для которых модулем является двойное отношение прямых — компонент множества нулей. Владимир Игоревич получил такую классификацию. Оказалось, что она включает три выделенных особенности, названные им параболическими (теперь они обычно называются простыми эллиптическими: simple elliptic), одну бесконечную трёхиндексную серию и ещё 14 особенностей, названных исключительными. (Параболические особенности встречались и раньше как первые особенности, которые «окружают» простые.) Возникла естественная проблема «придать смысл этой классификации», т. е. найти соответствующие ей известные математические объекты. Это оказалось очень трудной задачей, решённой Арнольдом только частично.

Владимир Игоревич составил огромную таблицу, в которую заносил всё, что было известно и становилось известно про эти 14 исключительных особенностей, и упорно искал в ней какие-то связи или намёки на них. В какой-то момент в таблице оказались тройки чисел, соответствующие этим исключительным особенностям и имеющие совершенно различную природу. Одна тройка чисел была определена И. Долгачёвым. Он заметил, что эти особенности соответствуют некоторым треугольникам на плоскости Лобачевского с углами π/α_1, π/α_2 и π/α_3, где α_1, α_2 и α_3 — натуральные числа. (Эти 14 треугольников имеют собственное независимое описание.) Тройку чисел $(\alpha_1, \alpha_2, \alpha_3)$ называют числами Долгачёва особенности.

С другой стороны, А. Габриэлов вычислял так называемые диаграммы Дынкина особенностей. Эти диаграммы описывают форму пересечения в группе двумерных гомологий слоя Милнора функции трёх переменных (а все унимодальные особенности можно считать функциями трёх переменных) в базисах специального вида. Эти базисы определены неоднозначно, и одной из целей Габриэлова было нахождение базисов, в которых диаграммы Дынкина выглядят проще всего: они содержат минимально возможное количество рёбер. Упрощение диаграмм требовало большой рутинной вычислительной работы (при отсутствии компьютеров) с плохо предсказуемым результатом: повезёт не повезёт, т. е. получишь «хорошую» диаграмму или не получишь. Да и хорошие могут быть, вообще говоря, разными. Габриэлову повезло (впрочем, везёт сильным): для 14 исключительных особенностей он получил похожие диаграммы. Они состояли из некоторого общего ядра, от которого отходили три «хвоста» длин γ_1, γ_2 и γ_3. Тройку чисел $(\gamma_1, \gamma_2, \gamma_3)$ называют числами Габриэлова особенности.

Происхождение этих чисел не имеет ничего общего, и довольно странно было бы сравнивать их друг с другом. Владимир Игоревич попробовал сделать это. (См. выше моё описание «немотивированных» действий Арнольда.) Как и следовало ожидать, для каждой исключительной особенности числа Долгачёва и числа Габриэлова по существу не имели ничего общего. Однако Арнольд заметил, что числа Долгачёва одной исключительной особенности совпадали с числами Габриэлова другой (иногда — той же) исключительной особенности, в то время как числа Долгачёва последней совпадали с числами Габриэлова первой. Тем самым на множестве исключительных особенностей возникала некоторая инволюция, меняющая между собой числа Долгачёва и Габриэлова.

Этот эффект Владимир Игоревич назвал странной двойственностью, которая в дальнейшем получила естественное имя странной двойственности Арнольда. (Для дальнейшего стоит упомянуть, что описанные события относятся к 1973–1975 гг.) Некоторое объяснение этой двойственности было дано Г. Пинкхамом (Henry Pinkham) в 1977 г. Оказалось, что числа Габриэлова особенности являются в некотором смысле числами Долгачёва на бесконечности для некоторой так называемой K3-поверхности, содержащей эту особенность, и наоборот.

Однако по большому счёту странная двойственность Арнольда оставалась некоторым курьёзом вплоть до конца века. В начале девяностых годов математики обнаружили, что физики изучают некоторые пары алгебраических многообразий с симметричными характеристиками. Такие многообразия назвали зеркально симметричными. Вначале происхождение таких многообразий (как и «доказательства» физиков того, что они имеют симметричные характеристики) выглядело для математиков некоторой мистикой. Однако позже соответствующие объекты и конструкции получили математическое описание. Как часто бывает, обнаружилось, что некоторые из них уже встречались в математике в другом виде. В частности, оказалось, что странная двойственность Арнольда — частный случай зеркальной симметрии. Сейчас признано, что странная двойственность Арнольда была первым в математике наблюдением эффекта зеркальной симметрии.

Каждый год (а часто и каждый семестр) семинар Арнольда начинался с занятия (иногда и двух), на котором он предлагал список открытых проблем. (Собрание этих проблем можно найти в книге: *Арнольд В. И.* «Задачи Арнольда», М.: Фазис, 2000. Многие из них не решены до сих пор.) Можно сказать, что в некотором смысле это были самые важные занятия семинара. На них участники семинара часто находили задачи для своей работы (иногда на многие годы вперёд).

Классификация критических точек функций в большинстве случаев требует анализа мономов «младших степеней», присутствующих в разложении Тейлора. Формализацией набора мономов младших степеней функции нескольких переменных является так называемая диаграмма Ньютона функции. Для многочлена от нескольких переменных диаграмма Ньютона описывает набор «крайних» мономов, присутствующих в многочлене. В 1968 году Арнольд сформулировал задачу: «Какие топологические характеристики вещественного (комплексного) многочлена вычислимы по диаграмме Ньютона (и знакам коэффициентов)?» В тот момент на неё «решателей» не нашлось. В 1975 году Владимир Игоревич сформулировал аналогичную задачу в локальной постановке: для критических точек функций. В том же году выражение числа Милнора в терминах диаграммы Ньютона было найдено А. Г. Кушниренко и опубликовано им в журнале Inventiones Mathematicae в 1976 году. Значение этой работы трудно переоценить. Её влияние на дальнейшие исследования можно продемонстрировать тем фактом, что в «MathSciNet» (онлайн-версии американского реферативного журнала «Mathematical Reviews») на неё зарегистрировано 203 ссылки. (Далеко не все математики имеют столько ссылок на все свои работы вместе взятые.) За ней последовали работы А. Г. Хованского и Д. Н. Бернштейна, описывающие другие инварианты критических точек и многочленов. Была понята связь этих задач с торической алгебраической геометрией, в которую в этот момент вдохнули новую жизнь. Так задача Арнольда дала старт бурному развитию одной из областей математики. Подобное происходило и во многих других областях.

Уже семь лет Владимира Игоревича нет с нами. Его ученики явно ощущают нехватку его вдохновляющего и направляющего внимания.

Сабир Меджидович Гусейн-Заде, мехмат МГУ, НМУ
sabirg@list.ru