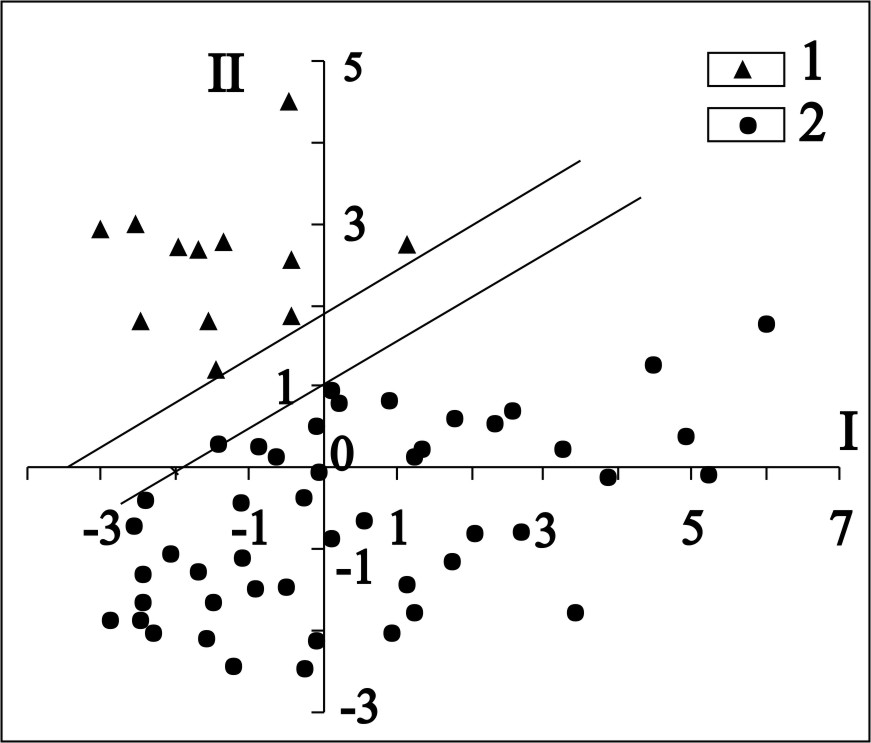
Пошаговый алгоритм 3 выполнения компонентного анализа

1. Ввод исходной матрицы **X**[*n*×*m*], где *n –* число объектов (точек наблюдения), *m –* число признаков (характеристик объекта)
2. Нормирование исходной матрицы – для каждого столбца вычисляется среднее значение и дисперсия; из каждого значения признака вычитается его среднее значение и делится на корень квадратный из дисперсии. Получается нормированная матрица исходных данных **Y**[*n*×*m*]. (Матрица выводится на экран и сохраняется в XL).
3. Вычисляется матрица коэффициентов корреляции **R**[*m*×*m*] для исходной матрицы (для нормированных значений – это просто перемножение двух векторов - признаков). (Матрица **R**[*m*×*m*] выводится на экран и сохраняется в XL).
4. Подключается специальная программа для расчета собственных значений и собственных векторов корреляционной матрицы (из справочника).
5. Матрица собственных векторов – это ортогональная матрица, столбцами которой являются собственные векторы матрицы коэффициентов корреляции, а матрица собственных значений – диагональная матрица, составленная из собственных значений (чисел) матрицы коэффициентов корреляции, соответствующих собственным векторам, причем элементы в диагональной матрице расположены в порядке убывания: 1 >2>...>m>0. (Матрицы выводятся на экран и сохраняются в XL).
6. После расчёта собственных значений и соответствующих им собственных векторов необходимо провести их упорядочение - рассчитанные собственные числа располагаются по мере уменьшения их значений - на первом месте располагается максимальное, затем самое большое из оставшихся собственных чисел и т.д. Последнее место занимает минимальное собственное значения. Поскольку каждому собственному значению соответствует свой собственный вектор, необходимо провести перестановку столбцов в матрице собственных векторов в том же порядке, как упорядоченные собственные числа. Сумма всех рассчитанных собственных значений должна равняться числу признаков в исходной матрице. Поскольку процедура нормирования переводит исходные переменные в безразмерные величины с нулевым средним и значением корня из единичной дисперсии каждого признака (а, следовательно, и стандартное отклонение тоже равно единице), поэтому суммарная дисперсия нормированных значений признаков равна количеству исходных признаков.
7. Точно так же сумма собственных чисел также равна числу признаков. И теперь можно рассчитать количество суммарной дисперсии для удовлетворения заданного заранее условия точности учета суммарной дисперсии – это сумма такого количества первых собственных чисел, которая и обеспечивает эту точность с учётом перевода в % при условии равенства *n* = 100%.
8. Матрица нагрузок главных компонент **A**[*m*×*q*] рассчитывается как произведение собственного вектора (из уже упорядоченной матрицы собственных векторов!!!!) на корень квадратный из соответствующего собственного числа, т.е. нагрузки на каждую ГК. Матрица выводится на экран и сохраняется в XL
9. Модель компонентного анализа предполагает точное определение, как компонентных нагрузок, так и значений главных компонент для каждой точки опробования (для каждого объекта). Матрица значений главных компонент находится по вычисленной предварительно матрице компонентных нагрузок и нормированной матрице исходных данных по следующему уравнению:

**F**[*n*×*q*] = **Y**[*n*×*m*]×**A**[*m*×*q*]×(**A**′[*q*×*m*]×**A**[*m*×*q*])-1

1. Матрица **F**[*n*×*q*] выводится на экран и сохраняется в XL).На этом этапе осуществляется проверка матрицы исходных данных на наличие смешанной совокупности. Для этого в XL на осях ГК1 и ГК2 строится облако точек для всех *n* объектов. Если облако точек распадается на две группы (см. рис.), это означает, что матрица исходных данных представляет собой смешанную совокупность, т.е. совокупность точек, отображаемых треугольниками и кружками, сформирована под воздействием различных природных (либо каких-либо других) процессов. В этом случае исходную матрицу **X**[*n*×*m*] следует разделить на две матрицы и обрабатывать их моделью КА отдельно.



1 – точки опробования глубоких водоносных горизонтов; 2 – скважины водоносных четвертичных и верхнеплиоценовых отложений (хоргосская свита) горизонтов

Рисунок В.1 – График распределения значений главных компонент для задачи А1 (смешанная совокупность)

Для удобства анализа на смешанную совокупность при построении облака точек желательно на рисунке указывать для каждой точки номер объекта (*n*).

1. В случае отсутствия смешанной совокупности процесс расчёта матрицы нагрузок и матрицы значений главных компонент можно считать завершённым и переходить к анализу точности вычислений.
2. Проверка точности вычислений матрицы нагрузок ГК и матрицы значений ГК осуществляется перемножением матриц **F** и **А** для получения восстановленной матрицы исходных данных в нормированном виде

**Y**восст[*n*×*m*] = **F**[*n*×*q*]×**A**′[*q*×*m*]

и визуального сравнения с рассчитанной ранее матрицей нормированных исходных данных **Y**[*n*×*m*].

1. Для облегчения визуальной оценки точности полученных решений восстановленные нормированные значения исходных данных переводятся в исходные единицы измерения умножением каждого элемента каждого столбца восстановленной нормированной матрицы на стандартное отклонение (корень из дисперсии) и суммированием со средним значением этого столбца.

Полученный результат восстановления исходных данных до исходных единиц измерения через матрицы **F** и **А** сохраняется в XL

**Для удобства анализа полученных результатов необходимо для каждой матрицы предусмотреть столбец и строку заголовка.**