

д.ф.-м.н. Вячеслав Иванович Кувыкин

**Анализ данных, поступающих
из систем измерения.
Согласование данных**

Материальный баланс



Примеры. Системы согласования баланса в промышленности (в нефтяной, пищевой, фармацевтической, распределение воды для снабжения городов, энергетика)

Под **материальным** балансом понимается концепция описания, основанная на законе сохранения масс, согласно которой извлечённая масса равна сумме изменения первоначальной и привнесённой в **динамическую** систему массы

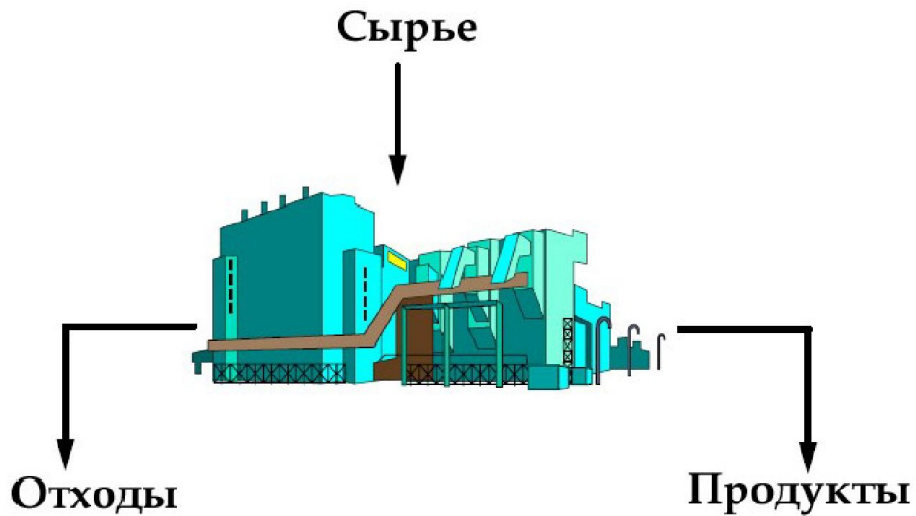
Лука Пачоли (годы жизни 1445г. – 1517г.) – математик из Италии, имя которого стало известно с момента обнаружения профессором математики Лючини в 1869 году в Миланском университете на забытых полках библиотеки известного на сегодняшний день каждому бухгалтеру Трактата о счетах и записях.

Ссылка на

сайт: <https://buhbook.net/buhgalterskij-uchet/istoriya-buhgalterskogo-ucheta/luca-pacholi-biografiya-osnovatelya-buhgalterskogo-ucheta-1445g-19-iyunya-1517g/>

Ж

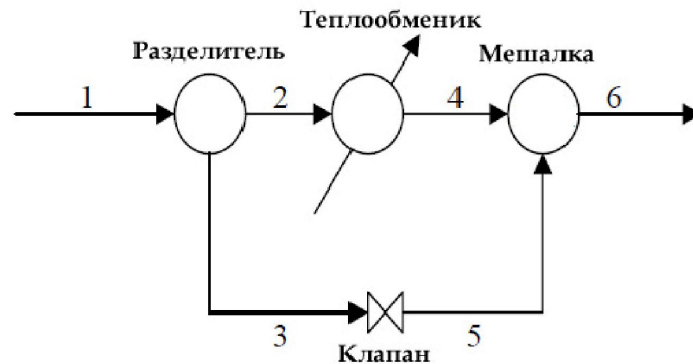
Составление уравнений массового баланса



Пример - водоснабжение, энергетика

Погрешности измерений – з-н сохранения массы не выполняется

Пример составления уравнения баланса



$$F1 - F2 - F3 = 0 ; \quad (1)$$

$$F2 - F4 = 0;$$

$$F3 - F5 = 0;$$

$$F4 + F5 - F6 = 0;$$

$$A = \begin{matrix} 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{matrix} \quad (2)$$

Согласование (сведение) уравнений массового баланса

В англоязычной литературе термин “**data reconciliation**” переводится как согласование данных и используется для математического обозначения термина сведение массового (материального) баланса. Основной целью согласования данных является вычисление так называемых истинных (согласованных) значений потоков и вычисление грубых ошибок измерений (“**gross error detection**”). Согласование производится на основе уравнений массового баланса и при этом должно соблюдаться правило минимума разницы (коррекции) между измеренными и согласованными величинами.

Пусть измеренное значение потока равно x_0 , а x будет истинное (реальное) значение потока, тогда:

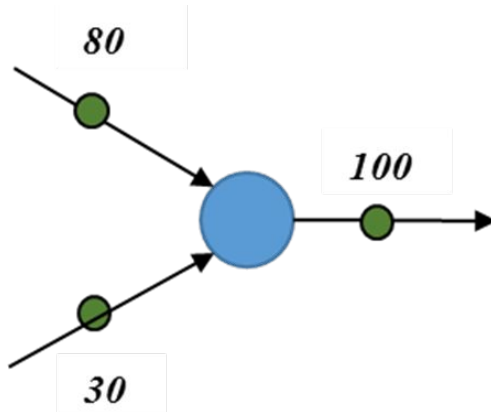
$$x_0 = x + \varepsilon, \quad (3)$$

где ε – ошибка измерения.

Согласно правилу минимума разницы (коррекции) между измеренными и согласованными величинами можно потребовать, чтобы разница между измеренными и согласованными величинами в квадрате была минимальной (метод наименьших квадратов)

Пример согласования данных, дебаланс

На вход поступил 2 потока, измеренная масса $x_{01} = 80$ и $x_{02} = 30$, на выходе $x_{03} = 100$.



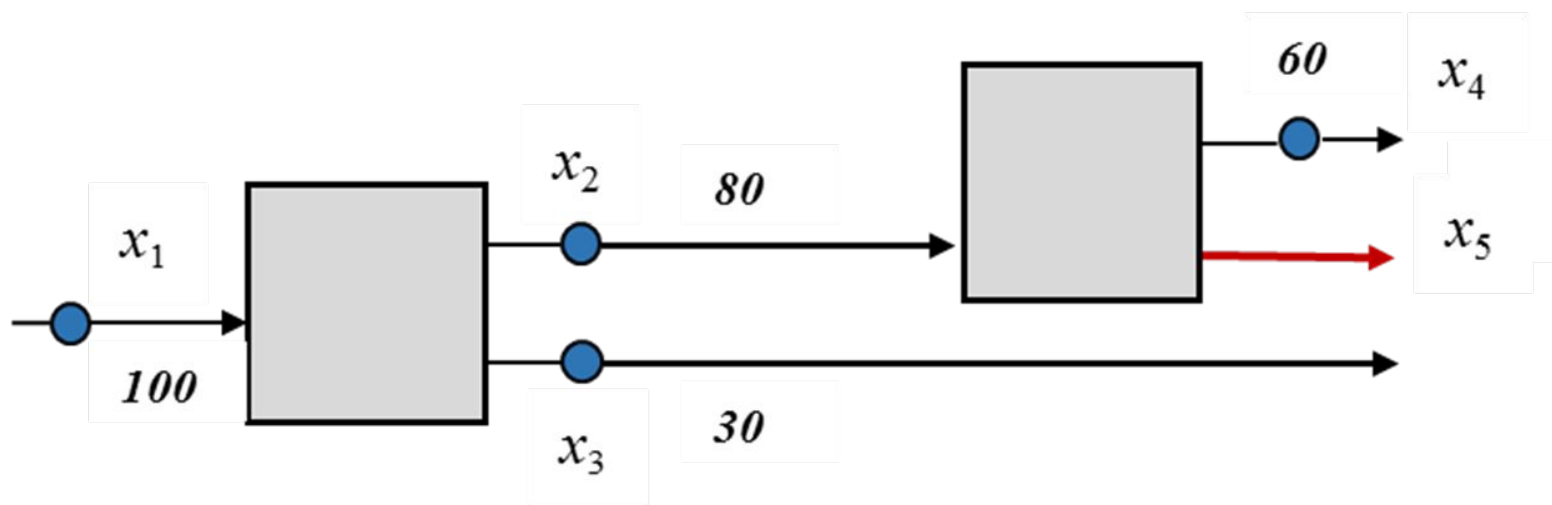
Уравнение материального баланса *вход-выход* для измеренных величин **не выполняется**, дебаланс $d=10$.

Требуется определить согласованные значения x , для которых выполняется закон сохранения масс и минимизируется некая целевая функция

СОГЛАСОВАНИЕ ДАННЫХ ПРИБОРНОГО УЧЕТА

1. Количество измерителей
2. Неизмеряемые потоки
3. Систематические и случайные погрешности
4. Диагностика неисправностей
5. Материальный и энергетический балансы

Типы потоков



Измеряемые, наблюдаемые
Показаны измеренные значения

Типы потоков

Обозначим подмножество **измеряемых потоков** $M^{(1)}$, **наблюдаемых** - $M^{(2)}$ и ненаблюдаемых - $M^{(3)}$, с элементами $x_k^{(1)} \in M^{(1)}$, $x_j^{(2)} \in M^{(2)}$, $x_i^{(3)} \in M^{(3)}$.

Вычислим минимальное количество k измерительных средств при отсутствии ненаблюдаемых потоков, т.е. множество $M^{(3)} = \emptyset$ - пустое.

Обозначим ранг матрицы $\text{rang}(\mathbf{A}) = r$. Как следует из решения системы линейных однородных уравнений, $n - r$ неизвестных могут быть выбраны произвольно и все решения системы могут быть получены построением линейной комбинации из $n - r$ независимых решений [9]. Таким образом, достаточное для определения всех значений расхода количество приборов $k = n - r$.

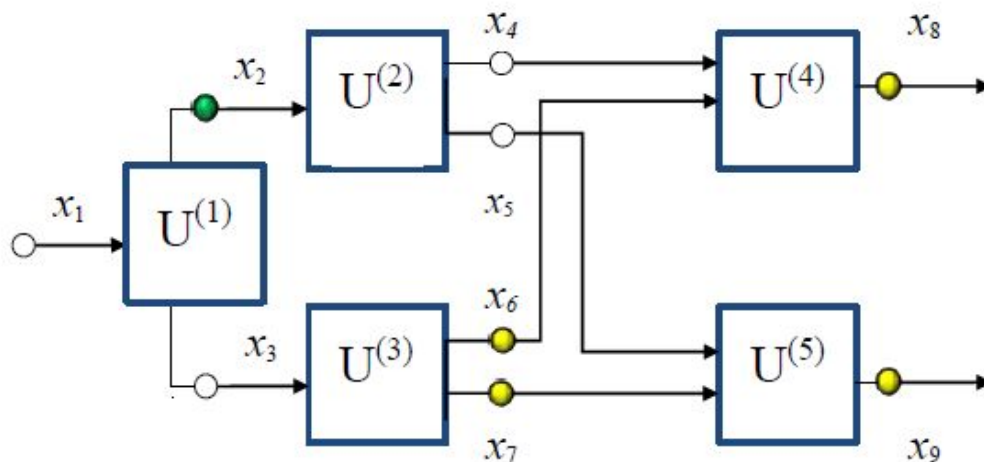
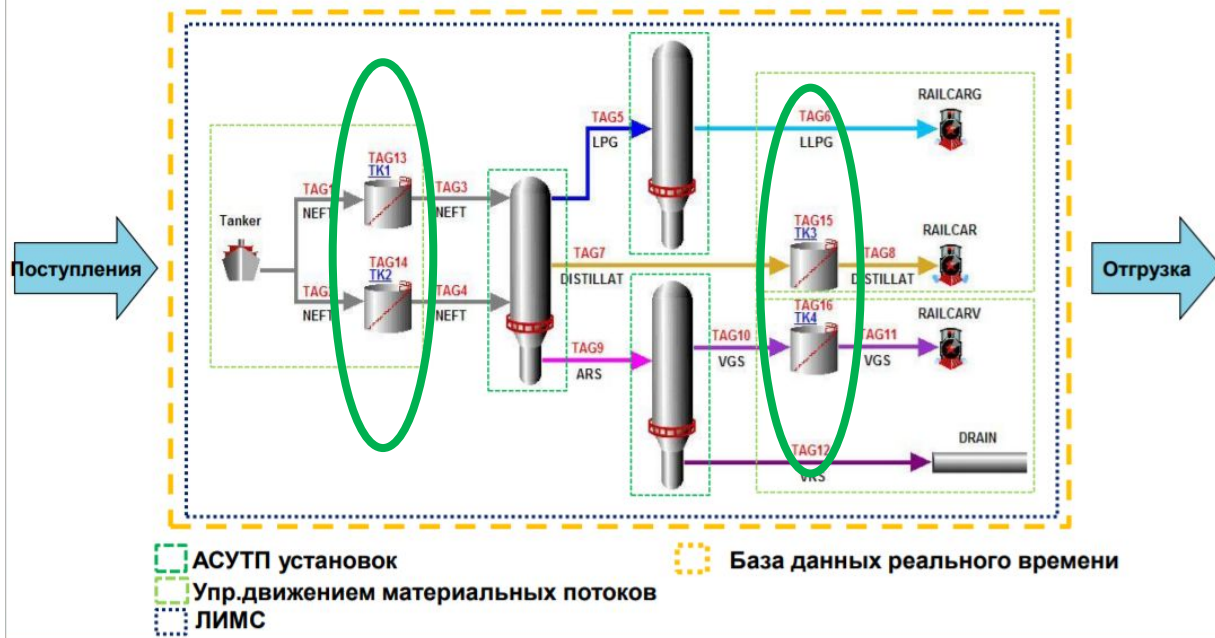


Рис. 1 – Схема потоков с расходом x_i с узлами $U^{(j)}$; x_6, x_7, x_8, x_9 - измеряемые значения, x_2 - дополнительное измеряемое значение, обеспечивающее учет при выходе из строя любого из приборов

Практический пример. Нефтепереработка

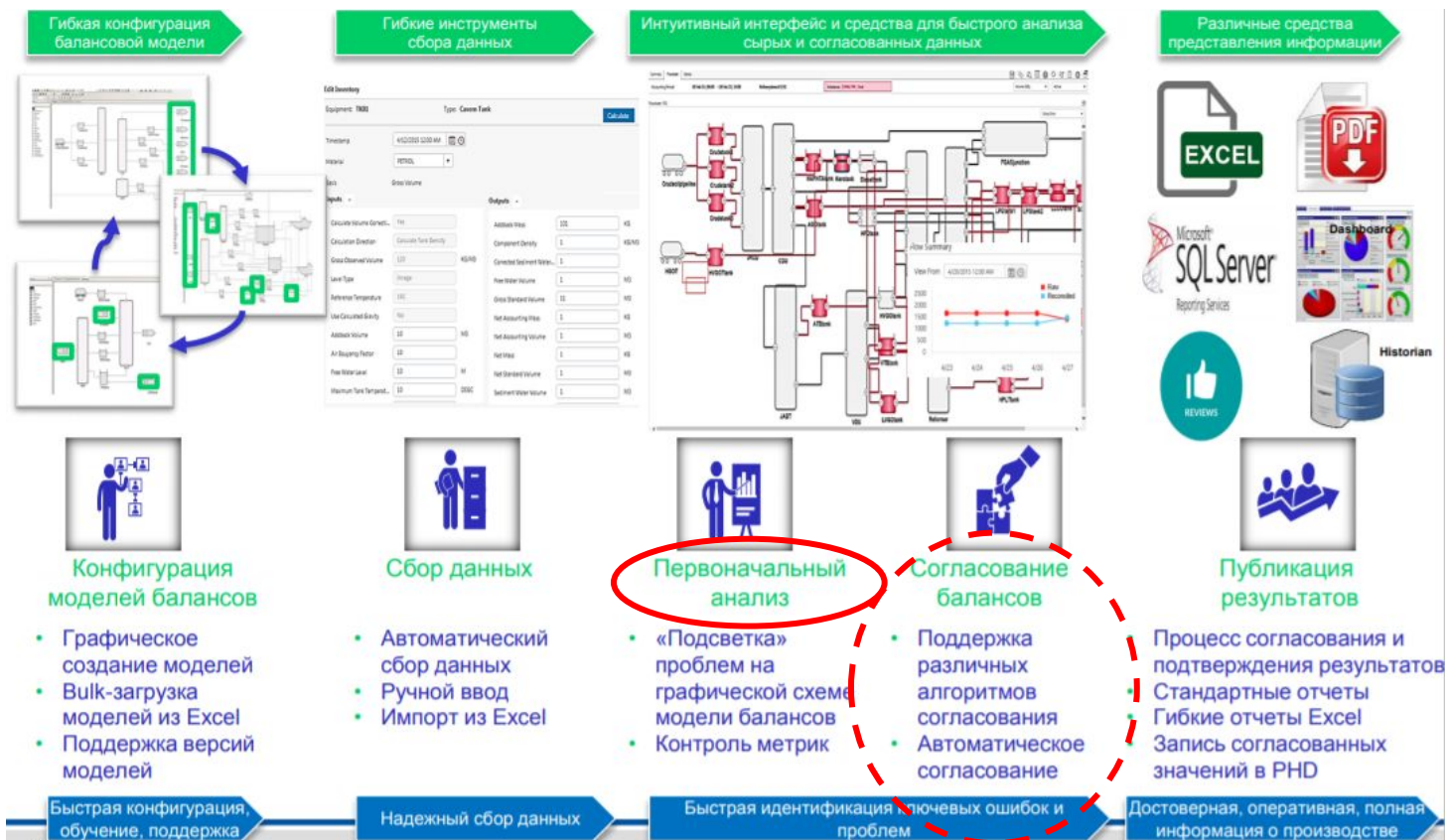
Исходные «сырые» данные поступают из разнородных СИСТЕМ



Производство и потребление по продуктам за период балансировки
Состояние запасов по продуктам на момент последнего закрытия баланса

Проблемы: Ошибочные измерения

Порядок действий при согласовании данных баланса



Математическая статистика. Обработка результатов измерений

Погрешности измерения

- **погрешность измерения:** разность между результатом измерения величины и действительным (опорным) значением величины;
- **случайная погрешность измерения,** случайная погрешность: составляющая погрешности измерения, изменяющаяся случайным образом (по знаку и значению) при повторных измерениях одной и той же величины, проведенных с одинаковой тщательностью;
- **систематическая погрешность измерения;** систематическая погрешность: составляющая погрешности измерения, остающаяся постоянной или закономерно изменяющаяся при повторных измерениях одной и той же величины, проведенных с одинаковой тщательностью;
- **грубая погрешность измерения:** погрешность измерения, существенно превышающая зависящие от объективных условий измерений значения систематической и случайной погрешностей.

Математическая модель согласования данных

$$\mathbf{Ax} = 0 \quad , \quad (1)$$

где \mathbf{x} – вектор согласованных переменных, описывающий массу материальных потоков предприятия, продукцию в резервуарах, расход топлива и потери, \mathbf{A} - матрица балансовых уравнений размерности $m \times n$ (элементы матрицы $a_{ij}=1$ для входящих потоков и $a_{ij}=-1$ - для выходящих, $a_{ij}=0$ - если поток не участвует в массовом балансе).

В качестве критерия оптимизации рассматривается функция $I(\mathbf{x})$, которая записывается в следующем виде

$$\bar{\mathbf{x}} = \arg \min \{ I(\bar{\mathbf{x}}) = (\bar{\mathbf{x}} - \mathbf{x}_0)^T \mathbf{B}^{-1} (\bar{\mathbf{x}} - \mathbf{x}_0) \mid \bar{\mathbf{x}} \in \mathbf{R}^q \} \quad (2)$$

где \mathbf{B} - матрица весовых коэффициентов, характеризующих погрешности соответствующих измерений, $\mathbf{B} = \text{diag}[\sigma_1^2, \sigma_2^2, \dots, \sigma_q^2]$ σ_q^2 - дисперсия, q – количество измерений, $\bar{\mathbf{x}}$ - согласованные переменные для измеренных значений.

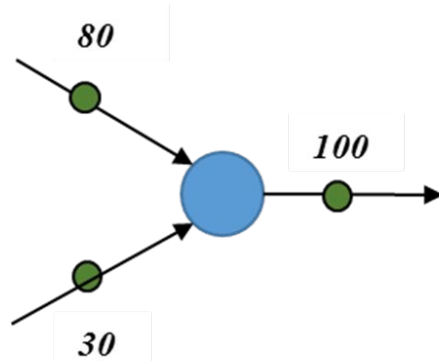
Физический смысл критерия (2) заключается в том, что решения системы уравнений (1), которые называются согласованными, должны как можно меньше отличаться от измеренных значений.

m – кол-во уравнений, n – кол-во переменных, q – кол-во измерений

Пример согласования данных с одним узлом

Проиллюстрируем работу алгоритма согласования данных на примере с одним узлом

На вход поступил 2 потока, измеренная масса $x_{01} = 80$ и $x_{02} = 30$, на выходе $x_{03} = 100$. Точность измерения одинакова



Для согласованных значений x выполняется закон сохранения масс и минимизируется целевая функция

$$l_{\min}(x) = \sum_{j=1}^3 \left(\frac{x_j - x_{0j}}{\sigma} \right)^2 \rightarrow \min$$

Если $\sigma = \varepsilon_j x_{0j}$ $l_{\min}(x) = \sum_j^k \left(\frac{x_j - x_{0j}}{\varepsilon_j x_{0j}} \right)^2 \rightarrow \min$

Здесь ε_j - погрешности соответствующих измерений

(Introduction to Gross Error Detection)

В некоторых случаях в измерениях может присутствовать так называемая грубая ошибка (gross error), которая возникает в случае смещения нуля, дрейфа значений вниз или вверх, а так же ухудшения точности.

*Грубые ошибки отрицательно влияют на согласование данных и для точного согласования данных необходимо вначале, **перед согласованием**, устранить все грубые ошибки измерений, а потом уже согласовывать данные.*

Закон больших чисел. Правило корня из n

Если надо измерить некоторую величину, истинное значение которой равно a , проводят n независимых измерений этой величины. В качестве меры истинного значения случайной величины выбирается средняя арифметическая. Если при измерениях отсутствуют систематические погрешности (искажающие результат измерения в одну и ту же сторону) естественно предположить, что математическое ожидание $M(\xi_i)=a$ при любом i . На основании теоремы Чебышева средняя арифметическая результатов n измерений сходится по вероятности к истинному значению a .

Если все измерения производятся с одинаковой точностью, характеризуемой дисперсией $D(\xi_i)=\sigma^2$, то дисперсия их средней

$$D\left(\frac{\sum_i \xi_i}{n}\right) = \frac{1}{n^2} D\left(\sum_i \xi_i\right) = \frac{1}{n^2} \sum_i D(\xi_i) = \frac{1}{n^2} (n\sigma^2) = \frac{\sigma^2}{n}$$

Таким образом, среднее квадратическое отклонение оказывается равным σ / \sqrt{n}

Средний ожидаемый разброс средней из n измерений в корень квадратный из n раз меньше разброса каждого измерения (этот вывод известен под названием «правила корня из n »)

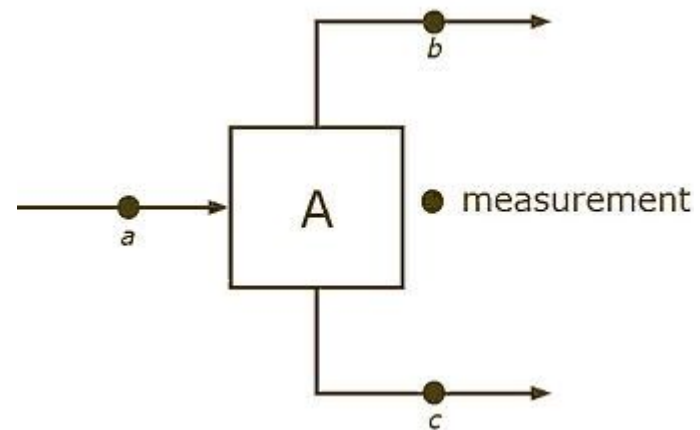
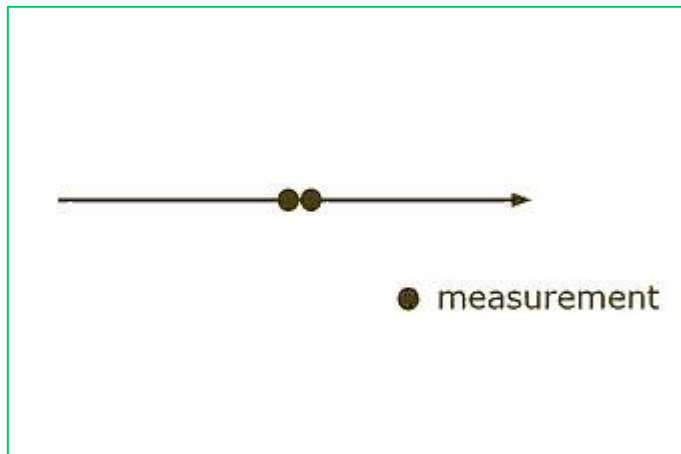


Замечание по точности измерения

Замечание. Если измерительный прибор имеет точность q (например, q - половина ширины деления шкалы прибора), то полученным способом нельзя получить точность измерения величины a , больше, чем q . Каждое измерение дает результат с неопределенностью q и среднее арифметическое будет обладать той же неопределённостью. (Н.Ш.Кремер)

Избыточные измерения

Избыточность может быть обусловлена **избыточностью датчиков**, когда датчики дублируются для того, чтобы иметь более одного измерения одной и той же величины. Избыточность также возникает, когда одна переменная может быть оценена несколькими независимыми способами из отдельных наборов измерений в данный момент времени или период усреднения по времени, используя алгебраические ограничения.



Наблюдаемость

Избыточность связана с понятием **наблюдаемости**. Переменная (или система) является наблюдаемой, если модели и измерения датчиков могут быть использованы для однозначного определения ее значения (состояния системы). Датчик является избыточным, если его удаление не приводит к потере наблюдаемости.

Избыточность может быть использована в качестве источника информации для перекрестной проверки и коррекции измерений и повышения их точности и точности: с одной стороны, они выверяются далее, проблема выверки данных, представленная выше, также включает в себя неизмеряемые переменные. Основываясь на информационной избыточности, оценки этих неизмеримых переменных могут быть рассчитаны вместе с их точностью. В промышленных процессах эти неизмеримые переменные, которые обеспечивает согласование данных, называются *мягкими* или виртуальными датчиками (soft sensors or virtual sensors), где аппаратные датчики не установлены.

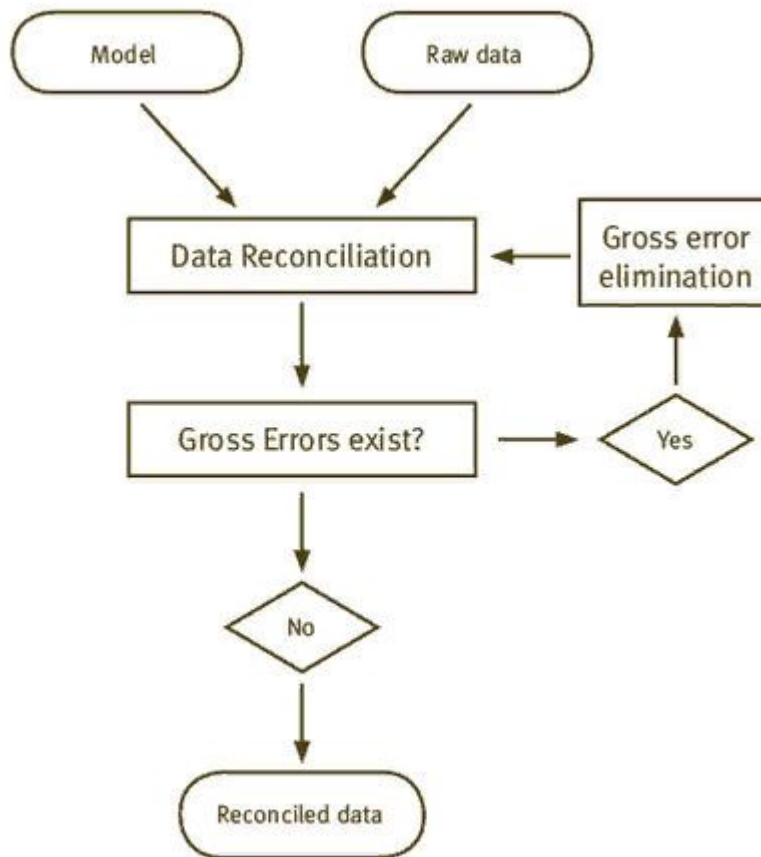
СОГЛАСОВАНИЕ ДАННЫХ

Data validation and reconciliation

Валидация и согласование данных промышленных процессов, или, более кратко, **согласования данных (process data reconciliation PDR)**- это технология, которая использует информацию о процессах и математические методы для автоматического обеспечения валидации и согласования данных путем коррекции измерений в промышленных процессах. Использование PDR позволяет извлекать точную и достоверную информацию о состоянии промышленных процессов из исходных данных измерений и получать единый согласованный набор данных, представляющих наиболее вероятную технологическую операцию.

Рабочий процесс проверки и согласования данных

Data validation and reconciliation



Методы решения

Квадратичное программирование (*quadratic programming, QP*) — задача оптимизации (минимизации или максимизации) **квадратичной функции** нескольких переменных при линейных ограничениях на эти переменные. Квадратичное программирование является частным случаем **нелинейного программирования**.

Ограничения – равенства

Задача квадратичного программирования несколько проще, если Q является положительно определённой и все ограничения являются равенствами (нет неравенств), поскольку в этом случае можно свести задачу к решению системы линейных уравнений.

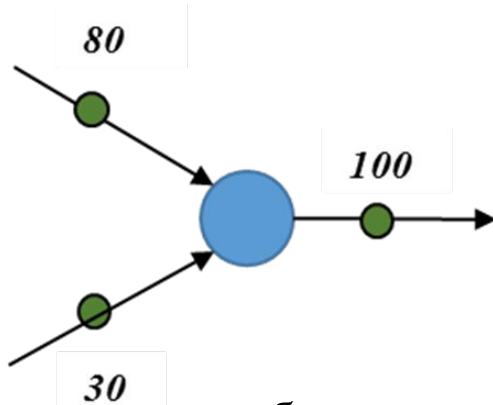
Алгоритм последовательного квадратичного программирования

Последовательное квадратичное программирование (*Sequential quadratic programming* (SQP)) — один из наиболее распространённых и эффективных оптимизационных **алгоритмов** общего назначения, основной идеей которого является последовательное решение задач квадратичного программирования, аппроксимирующих данную задачу **оптимизации**. Для оптимизационных задач без **ограничений** алгоритм SQP преобразуется в метод Ньютона поиска точки, в которой градиент целевой функции обращается в ноль. Для решения исходной задачи с ограничениями-равенствами метод SQP преобразуется в специальную реализацию ньютоновских методов решения системы Лагранжа.

Пример согласования данных с одним узлом

Проиллюстрируем работу алгоритма согласования данных на примере с одним узлом

На вход поступил 2 потока, измеренная масса $x_{01} = 80$ и $x_{02} = 30$, на выходе $x_{03} = 100$.



Уравнение материального баланса *вход-выход* для измеренных величин не выполняется, дебаланс $d=10$.

Для согласованных значений x выполняется закон сохранения масс

$$x_1 + x_2 - x_3 = 0$$

и минимизируется целевая функция

$$l_{\min}(x) = \sum_{j=1}^3 \left(\frac{x_j - x_{0j}}{\varepsilon_j x_{0j}} \right)^2 \rightarrow \min$$

Здесь ε_j - погрешности соответствующих измерений

Скрипт. Согласование баланса

```
# согласование баланса
#from scipy import optimize
import numpy as np
from scipy.optimize import minimize
import matplotlib.pyplot as plt
y=np.array([80,30,100]) #приборы
```

Библиотека `scipy.optimize`

`scipy.optimize.minimize`

`scipy.optimize.minimize`(*fun*, *x0*, *method=None*, *bounds=None*, *constraints=()*, *options=None*)

Минимизация скалярной функции одной или нескольких переменных.

fun Целевая функция, которая должна быть минимизирована.

`fun(x, *args) -> float`

где *x*- 1-D массив и *args* - кортеж фиксированных параметров, необходимых для полного задания функции.

x0 - *ndarray*, *shape* (*n*),

Первоначальные предполагаемые значения. Массив вещественных элементов размера (*n*), где 'n' - число независимых переменных.

bounds, *необязательно*

Границы переменных для методов SLSQP. Есть два способа указать границы:

Экземпляр [Bounds](#) класса.

Последовательность (*min*, *max*) пар для каждого элемента в *x*. *None* используется для указания отсутствия привязки.

constraints *ограничения*{*Constraint*, *dict*} или *список* {*Constraint*, *dict*}, *необязательный*

Определение ограничений.

```
from scipy.optimize import minimize
```

```
res = minimize(value, x0, method='SLSQP', constraints=cons,  
              bounds=bnds, options={'maxiter': 1000, 'ftol': 1e-12})
```



Метод SLSQP

`minimize(method='SLSQP')`

Метод **SLSQP** Sequential Least Squares Programming использует Последовательное квадратичное программирование для минимизации функции нескольких переменных с любой комбинацией ограничений границ, равенства и неравенства .

Метод включает в себя подпрограмму оптимизации SLSQP, первоначально реализованную Дитером Крафтом (Kraft, D. A software package for sequential quadratic programming. 1988. Tech. Rep. DFVLR-FB 88-28, DLR German Aerospace Center – Institute for Flight Mechanics, Koln, Germany). Обратите внимание, что оболочка обрабатывает бесконечные значения в границах, преобразуя их в большие плавающие значения.

Уравнения и ограничения

constraints ограничения {*Constraint, dict*} или список {*Constraint, dict*}, необязательный

Ограничения для SLSQP определяются как список словарей. Каждый словарь с полями:

- тип *str*
- Тип ограничения: 'eq' для равенства, 'ineq' для неравенства.
- **fun**
- Функция, определяющая ограничение.

```
#целевая ф-ция
```

```
def value(x):
```

```
    return ((y[0]-x[0])/y[0])**2 \
           + ((y[1]-x[1])/y[1])**2 + ((y[2]-x[2])/y[2])**2
```

```
#уравнения и ограничения
```

```
cons = {'type': 'eq',
```

```
        'fun': lambda x:
```

```
            np.array([x [0] + x [1]-x [2]
                      ])
```

```
bnds = ([70, 90], [20, 30], [100, 120]) #границы изменения
```

```
x0 = np.array([80, 30, 100]) # начальные значения
```



Минимизация

```
res = minimize(value, x0, method='SLSQP', constraints=cons,  
              bounds=bnds, options={'maxiter': 1000, 'ftol': 1e-12})  
print(res)  
print(res.message)
```

```
    fun: 0.005924657534246574  
    jac: array([-0.001, -0.001,  0.001])  
message: 'Optimization terminated successfully.'  
  nfev: 40  
   nit:  8  
  njev:  8  
status: 0  
success: True  
       x: array([ 75.616,  29.384, 105.   ])  
message  
Optimization terminated successfully.
```

Печать результатов

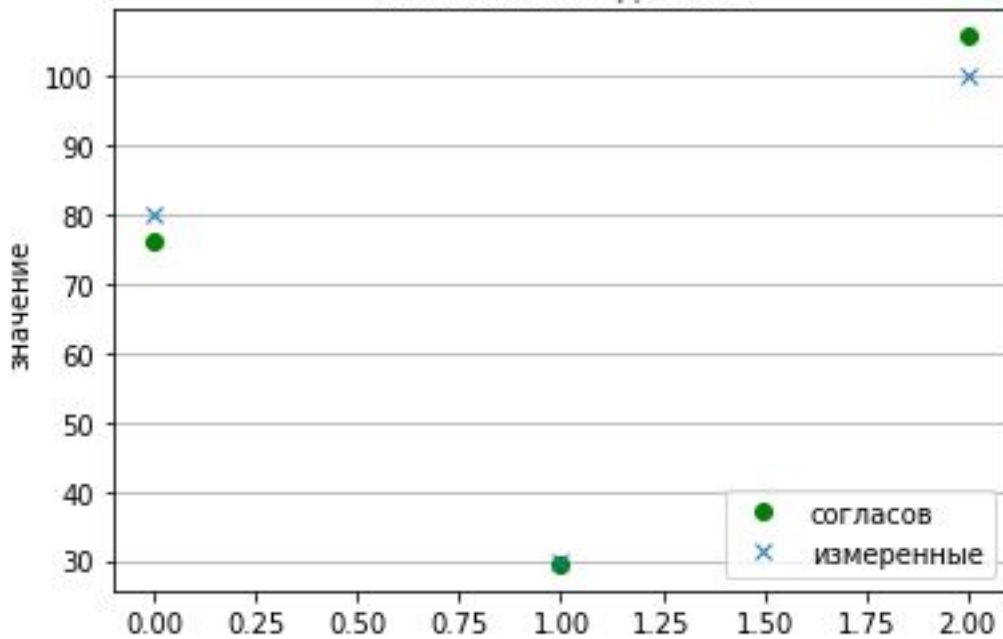
```
# красивая печать
np.set_printoptions(precision=3)
print('N измеренные согласованные коррекция')
for i in range(3):
    print('{: >7.3f} {:12.3f} {:9.3f}'.format(i+1, y[i],
        res.x[i], res.x[i]-y[i]))
```


Визуализация решения

```
#рисунок: измеренные и согласованные значения
x=np.arange(0,3,1)
plt.plot(x,res.x,'o',c='g',label='согласов')
plt.plot(x,y,'x',label='измеренные')
plt.grid(axis='y')
plt.legend(loc='lower right')
plt.title('согласование данных')
plt.ylabel('значение')
plt.show()
```

Согласованные данные (Python)

согласование данных



N измеренные согласованные коррекция

1	80.000	76.301	-3.699
2	30.000	29.480	-0.520
3	100.000	105.780	5.780

Задание 13_2 согласования данных с одним узлом

Проиллюстрировать работу алгоритма согласования данных на примере с одним узлом

На вход поступил 2 потока, измеренная масса за балансовый период

Вариант 1 $x_{01} = 150$ и $x_{02} = 40$;

Вариант 2 $x_{01} = 150$ и $x_{02} = 60$;

на выходе $x_{03} = 200$.

Для согласованных значений x выполняется закон сохранения масс и минимизируется целевая функция

$$l_{\min}(x) = \sum_{j=1}^3 \left(\frac{x_j - x_{0j}}{\varepsilon_j x_{0j}} \right)^2 \rightarrow \min$$

Здесь ε_j - погрешности соответствующих измерений принять равной единице

1. Написать уравнение баланса
2. Вычислить согласованные значения с использованием Python
3. Ответить на вопрос «Имеются ли грубые ошибки измерения?» и обосновать свой ответ.

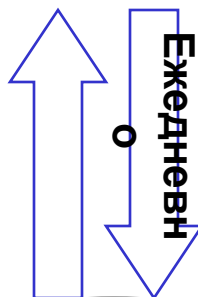
Инженерные приложения. Согласование данных

Схема работы IT-системы Production Balance

Production Balance (Honeywell)

Input Units (BBL)					
Inventory					
	Open	Close	Receipts	Shipments	Imbalance
Raw Balance	3603556.26	4010146.40	548854.16	133647.46	8616.57
Reconciled Balance	3603556.26	4020253.82	548854.16	133647.46	-1490.85
Adjustment		10107.42		.00	

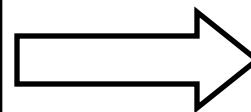
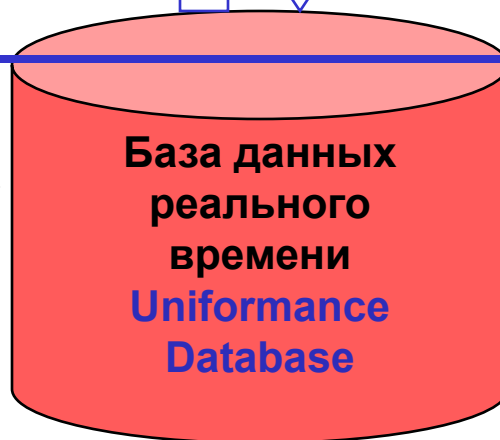
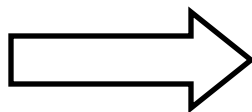
“Сырые” потоки
Запасы
Перемещения н/п



Сбалансированные потоки
Запасы
Перемещения
Статистика по балансам

Источники информации

Системы управления
Датчики в резервуарах



Другие приложения и ERP (SAP)

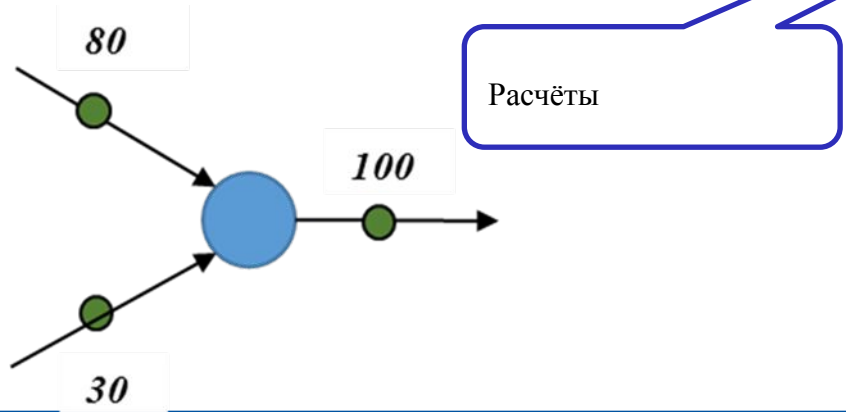
Пример практического использования системы

Проверка пределов

Линия	Вход (Изм.)	Нач. запас (Изм.)	Выход (Изм.)	Кон. запас (Изм.)	Вход (Согл.)	Нач. запас (Согл.)	Выход (Согл.)	Кон. запас (Согл.)	Ошибка
	110,000		100,000		105,780		105,780		1,204

Линия	Материал	Вкл. в расчет	Датчик	Тип	Изм.	Скорр.	Согл.	Коррекция	% корр.	Ошибка
	NEFT	<input checked="" type="checkbox"/>	RR.100012	Изм.	80,000	80,000	76,301	-3,699	-4,624	7,603
	NEFT	<input checked="" type="checkbox"/>	RR.100011	Изм.	30,000	30,000	29,480	-0,520	-1,734	7,603

Линия	Материал	Вкл. в расчет	Датчик	Тип	Изм.	Скорр.	Согл.	Коррекция	% корр.	Ошибка
	NEFT	<input checked="" type="checkbox"/>	RR.100013	Изм.	100,000	100,000	105,780	5,780	5,780	7,603



Типичный материальный баланс процесса каталитического крекинга гидроочищенного вакуумного газойля

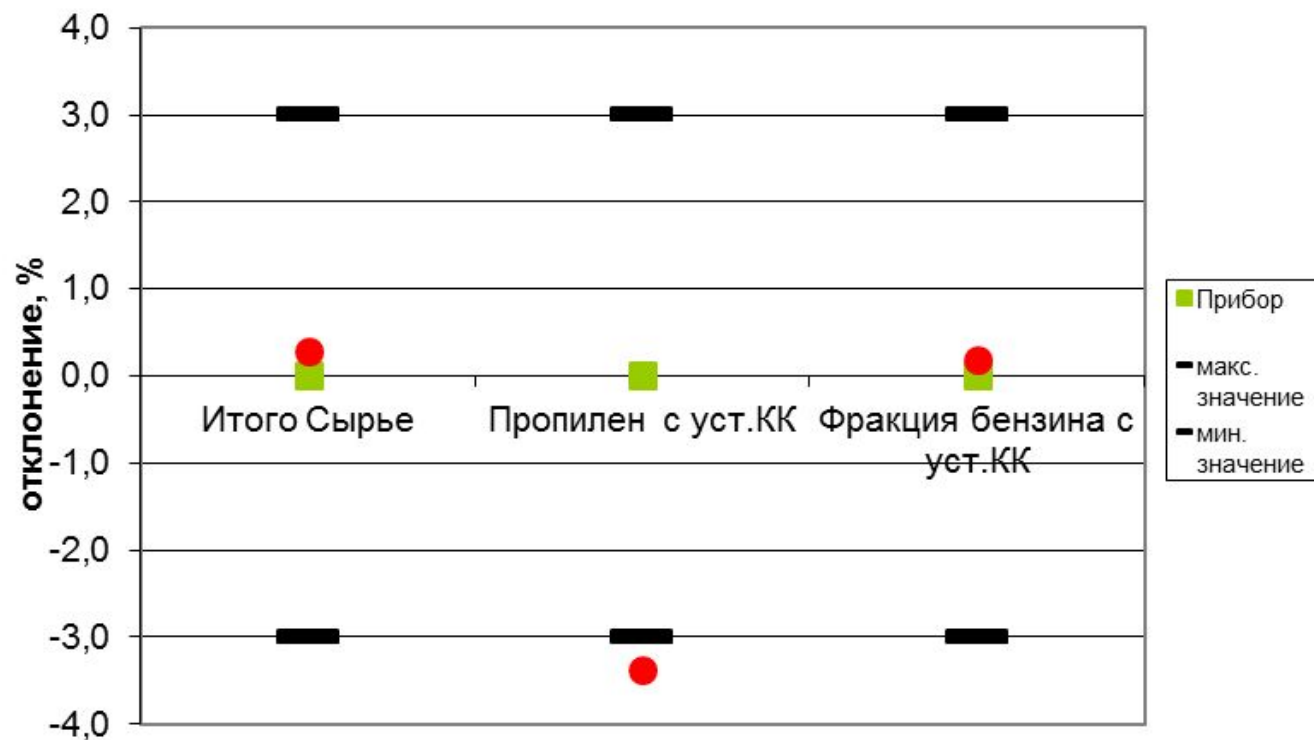
Продукция	Выход % на сырье
Взято всего:	100
Гидроочищенный вакуумный газойль (Фр.350-500°C)	100
Получено всего:	100
<u>H₂</u> (водород)	0,04
<u>CH₄</u> (метан)	0,25
<u>C₂H₆</u> (этан)	0,23
<u>C₂H₄</u> (этилен)	0,36
<u>C₃H₈</u> (пропан)	0,85
<u>C₃H₆</u> (пропилен)	2,73
<u>C₄H₁₀</u> (бутан)	0,89
<u>C₄H₈</u> (бутены)	2,5
<u>(CH₃)₃CH</u> (изобутан)	4,20
бензиновая фракция (ОЧИ-91/92)	58,62
газойль (легкий+тяжелый)	27,17
Кокс + потери	2,17

Инженерные приложения. Установка Каталитического Крекинга (УКК)

Scalpellum					
(условные данные)					
УКК	Прибор		Сбал. знач.		Откл. сбал. прибор(+,-)
	т.т	%	т.т	%	
Г/о фракция вакуумного газойля	79,551	100,00	79,780	100,00	0,229
Некондиция на УКК	0,000	0,00	0,000	0,00	0,000
Итого Сырье	79,551	100,00	79,780	100,00	0,229
Сухой газ	2,407	3,03	2,710	3,40	0,303
Фракция пропановая с уст.КК	1,058	1,33	1,058	1,33	0,000
Пропилен с уст.КК	6,158	7,74	5,950	7,46	-0,208 >3%
Фракция пропан-пропиленовая с уст. КК	0,000	0,00	1,007	1,26	1,007
Фракция бутан-бутиленовая с уст.КК	11,545	14,51	11,308	14,17	-0,237
Фракция бензина с уст.КК	43,630	54,85	43,715	54,79	0,085
Фракция легкого газойля с уст.КК	8,555	10,75	8,658	10,85	0,103
Фракция тяжелого остатка с уст.КК	1,440	1,81	1,508	1,89	0,068
Сероводород	0,015	0,02	0,015	0,02	0,000
Кокс	3,270	4,11	3,270	4,10	0,000
Потери	0,000	0,00	0,581	0,73	0,581
Итого Выработка	78,078	98,15	79,780	100,00	1,702
Дебаланс (поступление - выработка)	1,473		0,000		

Визуализация данных УКК

Сбалансированные данные (с начала месяца, дов. инт. 3%)



Вопросы ?

Спасибо за внимание !