требителю направляются те b_1 единиц продукта, смесь которых обеспечила равенство количества примеси; после этого снова остается задача меньшего размера, где выполняются условия (3) — (5). Оставшаяся задача меньшего размера решается аналогично, пока не будет распределен весь продукт, т. е. получено некоторое решение. Теорема доказана.

К сожалению, аналогичная простая теорема в задаче о транспортировке смесей более чем двух компонент не имеет места.

Поступила в редакцию
20 I 1968

О СИСТЕМЕ МАССОВОГО ОБСЛУЖИВАНИЯ, СОДЕРЖАЩЕЙ КАНАЛЫ С РАЗЛИЧНОЙ ДИСЦИПЛИНОЙ ОБСЛУЖИВАНИЯ

А. С. НЕМИРОВСКИЙ, Б. К. КАЛЯКИН, Л. Н. ЗАКАШАНСКИЙ

(Москва)

Встречаются системы массового обслуживания, включающие универсальные и специализированные каналы. Первые из них могут обслуживать требования различных типов, т. е. как обслуживаемые, так и не обслуживаемые на других — специализированных каналах системы. Специализированные же каналы принимают на обслуживание только однотипные требования [1].

Простейшая схема системы массового обслуживания, каналы которой отличаются друг от друга дисциплиной обслуживания, приведена на рис. 1, где I и II — соответственно специализированный и универсальный каналы; λ_1 и λ_2 — плотности входящих потоков требований, поступающих соответственно на I и II канал; λ_{12} и λ_{22} — плотности требований из первой и второй очередей, составляющих входящий поток для II канала.

Поступая в систему, требования первого и второго типа первоначально становятся в первую и вторую очередь соответственно.

Такого рода схемы систем массового обслуживания встречаются, например, в поверочных лабораториях, куда доставляются с предприятий на поверку измерительные приборы. Обслуживание осуществляется на поверочных установках, выполняющих в зависимости от их свойств роль специализированных или универсальных каналов.

Рациональное функционирование подобной системы, состоящее в достаточно высокой загрузке обслуживающих каналов при возможно меньшем среднем времени ожидания требований в очереди, достигается, кроме правильного выбора количества каналов, еще и целесообразной стратегией их загрузки.

Для систем массового обслуживания с ожиданием и каналами, имеющими одинаковую доступность, также могут быть приемлемы различные стратегии загрузки каналов. В частности, в [2] предложен ряд стратегий, направленных на минимизацию времени ожидания и рекомендуемых для выбора требованиями одной из двух очередей в условиях двухканальной системы с одинаковой дисциплиной обслуживания каналов, но с различными вариантами дисциплины ожидания.

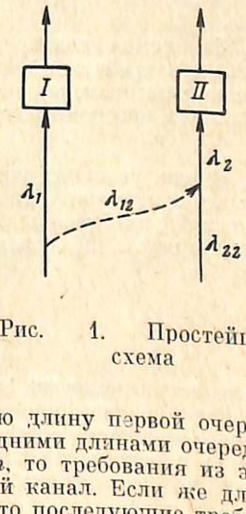

Рис. 1. Простейшая схема

В настоящей работе рассматривается задача выбора стратегии загрузки каналов системы, обеспечивающей либо равенство средних времен простоев каналов с одинаковой и различной дисциплиной обслуживания, либо равенство средних времен ожидания в очереди всех требований (в том числе разнотипных), либо минимизацию общих потерь от ожиданий и простоев. Поскольку в данном случае из-за наличия в системе каналов с различной дисциплиной обслуживания возможен переход только из первой очереди во вторую (см. рис. 1), указанная выше задача сводится к задаче определения наивыгоднейшего с той или иной точки зрения правила одностороннего перехода требований из очереди в очередь.

Представляется целесообразным ограничить наибольшую длину первой очереди так, чтобы установить определенное соотношение между средними длинами очередей. Положим, что если длина первой очереди не превышает m, то требования из этой очереди будут поступать на обслуживание только в первый канал. Если же длина первой очереди в некоторый момент становится равной m, то последующие требования, начиная с $(m+1)$-го, из первой очереди направляются во вторую, увеличивая плотность второго входящего потока.

Оба входящих потока предполагаются простейшими, а распределение времени обслуживания на обоих каналах — показательным.

Вследствие принятых предположений система уравнений, описывающих работу специализированного канала, будет той же, что и представленная, например, в [3] для одноканальной системы смешанного типа с ограничением по длине очереди.

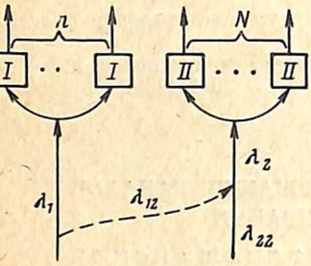

Рис. 2. Общая схема

Обозначив $\alpha_1 = \lambda_1 / \mu_1$, где μ_1 — величина, обратная среднему времени обслуживания одного требования в первом канале, получим вероятность того, что первый канал свободен

$$P_0 = \frac{1-\alpha_1}{1-\alpha_1^{m+2}}. \tag{1}$$

Вероятность отказа в обслуживании на специализированном канале для требований из первой очереди равна

$$P_{1+m} = \frac{\alpha_1^{m+1}(1-\alpha_1)}{1-\alpha_1^{m+2}}. \tag{2}$$

В практике работы некоторых поверочных лабораторий чаще, чем простейшая схема использования каналов с различной дисциплиной обслуживания, применяется схема, подобная представленной на рис. 2, при которой на N универсальных каналов приходится n специализированных. Тогда формулы вероятностей состояний специализированных каналов при $0 \leqslant k \leqslant n$ и при $1 < i \leqslant m$, согласно [3], принимают вид

$$P_k = \frac{\dfrac{\alpha_1^k}{k!}}{\sum\limits_{k=0}^{n} \dfrac{\alpha_1^k}{k!} + \dfrac{\alpha_1^n}{n!} \sum\limits_{i=1}^{m} \left(\dfrac{\alpha_1}{n}\right)^i}, \qquad P_{n+i} = \frac{\dfrac{\alpha_1^n}{n!}\left(\dfrac{\alpha_1}{n}\right)^i}{\sum\limits_{k=0}^{n} \dfrac{\alpha_1^k}{k!} + \dfrac{\alpha_1^n}{n!} \sum\limits_{i=1}^{m} \left(\dfrac{\alpha_1}{n}\right)^i}. \tag{3}$$

Средняя длина очереди из требований, поступающих на специализированные каналы, равна $l_1 = \sum\limits_{i=1}^{m} iP_{n+i}$, а в случае, когда такой канал один,

$$l_1 = \sum_{i=1}^{m} iP_{1+i} = \frac{\alpha_1^2}{(1-\alpha_1^{m+2})(1-\alpha_1)}\left[1-(m+1)\alpha_1^m + m\alpha_1^{m+1}\right]. \tag{4}$$

За промежуток времени $(t;\ t+\Delta t)$ из первой очереди во вторую направляется $\lambda_1 P_{n+m}\Delta t$ требований. Плотность потока переходов из первой очереди во вторую в сумме с потоком, характеризуемым λ_{22}, есть плотность входящего потока для второго канала, которая составляет

$$\lambda_2 = \lambda_{22} + \lambda_1 P_{n+m}. \tag{5}$$

В этих условиях универсальный канал можно рассматривать как систему с ожиданием и неограниченной по длине очередью. Если второй входящий поток обслуживается N универсальными каналами, то вероятность нахождения в системе k требований при $k \leqslant N$, согласно [4], равна

$$Q_k = \frac{\alpha_2^k}{N!} \cdot \left[\sum_{k=0}^{N} \frac{\alpha_2^k}{k!} + \frac{\alpha_2^{N+1}}{N!(N-\alpha_2)}\right]^{-1}, \tag{6}$$

где α_2 есть отношение λ_2 к μ_2 — средней пропускной способности второго канала.

Если средняя производительность этого канала при обслуживании требований первого и второго типа будет разной, то средние пропускные способности второго канала в отношении таких требований обозначим через μ_{12} и μ_{22} соответственно. В этих условиях формула (5) примет вид

$$\lambda_2 = \lambda_{22} + \lambda_1 P_{n+m}\mu_{12} / \mu_{22}. \tag{5а}$$

Нетрудно обнаружить, что в случае, когда универсальный канал один, вероятность его незагрузки составит

$$Q_0 = 1 - \alpha_2. \tag{7}$$

Таким же образом, рассматривая второй канал как одноканальную систему с ожиданием и неограниченной очередью, получим среднюю длину очереди, поступающей на этот канал

$$l_2 = \alpha_2 \cdot (1 - \alpha_2)^{-1}. \tag{8}$$

Наивыгоднейшее правило перехода, выражаемое выбранным значением m, может устанавливаться исходя из различных точек зрения.

Можно руководствоваться соображением равенства времени простоев специализированных и универсальных каналов, т. е. условием

$$\frac{1}{n}\sum_{k=0}^{n} P_k(n-k) = \frac{1}{N}\sum_{k=0}^{N} Q_k(N-k), \tag{9}$$

при котором обеспечивается равномерная загрузка всех каналов. При $n = 1$ левая часть уравнения приобретает вид P_0, а при $N = 1$ правая часть — Q_0; их выражения представлены в формулах (1) и (7).

Способ получения m для данной стратегии может быть проиллюстрирован на числовом примере.

Пример 1. Определить m, исходя из вышеуказанного соображения и из условия, что многоканальная система массового обслуживания содержит четыре специализированных и один универсальный канал; параметры поступления и обслуживания таковы: $\lambda_1 = 40$, $\mu_1 = 10$, $\lambda_{22} = 9$ и $\mu_2 = 12$ (требований в день).

Решение. Имеем $\alpha_1 = \lambda_1 / \mu_1 = 4$. По формулам (3) и (9) находим средний коэффициент простоев для каждого из четырех специализированных каналов

$$\frac{1}{4}\sum_{k=0}^{4} P_k(4-k) = \frac{10{,}7}{34{,}3 + 10{,}7m}.$$

Подставляя (5) в выражение $\alpha_2 = \lambda_2 / \mu_2$, получим $\alpha_2 = {}^3/_4 + 3^1/_3 P_{n+m}$.

Затем, пользуясь формулами (7) и (3) для P_{n+i}, вычислим $Q_0 = 0{,}25 - 35{,}7(34{,}3 + 10{,}7\,m)^{-1}$.

Приравнивая $\frac{1}{4}\sum_{k=0}^{4} P_k(4-k)$ к Q_0, получим уравнение

$$\frac{10{,}7}{34{,}3 + 10{,}7m} = 0{,}25 - \frac{35{,}7}{34{,}3 + 10{,}7m},$$

решив которое, находим, что $m = 14$.

Можно также исходить из условия равенства средних времен ожидания требований, поступающих на специализированные и на универсальные каналы. В этом случае должны быть такие средние длины очередей, чтобы, согласно [5], $l_1 / \lambda_1 = l_2 / \lambda_2$. Выбор m, отвечающего этому равенству, направлен на выполнение такой стратегии загрузки каналов, которая обеспечит одинаковое среднее время ожидания начала обслуживания для требований обоих типов. Числовой пример содержит способ нахождения m, величина которого обусловлена указанным соображением.

Пример 2. Найти m, исходя из равенства $l_1 / \lambda_1 = l_2 / \lambda_2$, при условии, что в системе один специализированный и один универсальный канал (см. рис. 1) и следующие параметры поступления и обслуживания: $\lambda_1 = 15$, $\lambda_{22} = 11$, $\mu_1 = \mu_2 = 17$ (требований в день).

Решение. Имеем $\alpha_1 = 0{,}9$. Далее, пользуясь формулой (4) для средней длины очереди у первого канала, определяем $l_1 / \lambda_1 = 5{,}4(1 - 0{,}1m0.9^m - 0{,}9^m)(10 - 8{,}1 \times 0{,}9^m)^{-1}$.

Из (5) следует, что $\alpha_2 = 0{,}65 + 0{,}88P_{1+m}$.

Отсюда, согласно (2), вычисляем $\alpha_2 = (6{,}5 - 4{,}4\cdot 0{,}9^m)(10 - 8{,}1\cdot 0{,}9^m)^{-1}$.

Подставляя значения α_2 и $1 - \alpha_2$ в (8), находим среднюю длину очереди у второго канала l_2, после чего, пользуясь (5), нетрудно получить необходимое выражение для l_2 / λ_2.

Приравнивая l_1 / λ_1 к l_2 / λ_2, получаем следующее уравнение

$$\frac{5{,}4(1 - 0{,}1m\cdot 0{,}9^m - 0{,}9^m)}{10 - 8{,}1\cdot 0{,}9^m} = \frac{(6{,}5 - 4{,}4\cdot 0{,}9^m)}{(3{,}6 - 3{,}7\cdot 0{,}9^m)(110 - 102.5\cdot 0{,}9^m)}.$$

Наиболее близким к решению этого уравнения целым значением m является $m = 4$.

Зная значения среднего времени ожидания для обоих каналов, нетрудно решить вопрос о степени рациональности применения схемы с универсальным каналом в сравнении со схемой без универсального канала. Применяя в условиях примера 2 правило одностороннего перехода из очереди в очередь, выраженное полученным значением $m = 4$, имеем $l_1 / \lambda_1 = l_2 / \lambda_2 = 0{,}10$ (дня).

В случае же применения схемы, в которой те же два канала работают как специализированные на разных типах требований, эти каналы должны рассматриваться как две отдельные системы с ожиданием и неограниченными очередями. В этих условиях среднее время ожидания для двух каналов равно $0{,}5(l_1' / \lambda_1 + l_2' / \lambda_{22})$, где l_1' и l_2' — средние длины очередей, поступающих соответственно на первый и второй специализированные каналы. Эти длины определяются по формуле (8) и для параметров поступления и обслуживания того же примера 2 составляют: $l_1' = 0{,}9(1 - 0{,}9)^{-1} = 9{,}0$; $l_2' = 0{,}65(1 - 0{,}65)^{-1} = 1{,}8$.

Таким образом, $0{,}5(l_1' / \lambda_1 + l_2' / \lambda_{22}) = 0{,}38$ (дня).

Приведенный пример показывает, насколько при обслуживании разнотипных требований рациональнее (с точки зрения значений среднего времени ожидания, которое иногда рассматривают как критерий эффективности системы) применение схемы с универсальными каналами в сравнении со схемой, содержащей только специализированные каналы. Согласно приведенным расчетам, в данном случае наличие универсального канала вместо второго специализированного значительно сокращает общее среднее время ожидания в очереди [0,10 против 0,38 дня].

В практике работы поверочных лабораторий значения времени ожидания и времени пребывания требований в системах массового обслуживания, как правило, значительно больше, чем для условий примера 2, так как на обслуживание в лабораторию обычно требования поступают не ординарным входящим потоком, а группами. Размер такой группы требований является случайной величиной с большой дисперсией. Время пребывания требований в системе при групповом поступлении всегда больше, хотя бы потому, что в данном случае каждое требование может считаться обслуженным только после завершения обслуживания всей группы. Способ приближенного вычисления среднего времени пребывания требований в системе в условиях группового входящего потока отражен в следующем примере.

Пример 3. Вычислить среднее время пребывания требований в одноканальной системе, для которой параметры прибытия и обслуживания требований подобны тем, что заданы для первого канала из примера 2 (для него $\lambda_1 = 15$ и $\mu_1 = 17$ требований в день, а среднее время ожидания, когда этот канал рассматривался отдельно, составляло 0,54 дня, что в сумме со средним временем обслуживания, равным $1 / \mu_1 = 0{,}06$ дня, дает время пребывания требований в системе средней продолжительностью 0,60 дня). Дополнительными условиями служат те обстоятельства, что требования поступают группами, средний размер которых $r = 15$ требованиям, средняя дневная интенсивность поступления групп $\psi = 0{,}72$, дисперсия размера группы $Д\ (r) = 191$ и дисперсия времени обслуживания одного требования $Д\ (1 / \mu_1) = 0{,}0024$.

Р е ш е н и е. Согласно результатам, изложенным в [6], при групповом поступлении среднее время ожидания первого из группы требований для данного случая составляет

$$T_{\text{ож}} = 0{,}5\varphi[(1 / \mu_1)_2 r + (1 / \mu_1)^2 (r_2 - r)](1 - \psi r / \mu_1)^{-2}, \tag{10}$$

где $(\mu_1^{-1})_2 = D(\mu_1^{-1}) + \mu_1^{-2} = 0{,}0024 + 0{,}0036 = 0{,}006$; $r_2 = D(r) + r^2 - r = 191 + 225 - 15 = 401$.

Тогда

$$T_{\text{ож}} = 0{,}36 \frac{0{,}006 \cdot 12{,}5 + 0{,}0036(401 - 15)}{\left(1 - \dfrac{0{,}72 \cdot 15}{17}\right)^2} = 4{,}1 \text{ (дня)}.$$

Среднее время обслуживания данной группы требований равно $T_{\text{об}} = r / \mu_1 = 0{,}9$ (дня).

Таким образом, среднее время пребывания требований в рассматриваемой системе есть $T_{\text{пр}} = T_{\text{ож}} + T_{\text{об}} = 5$ (дней).

Как указывалось выше, то или иное соображение при организации обслуживания учитывается соответствующим правилом перехода из очереди в очередь, которое через вычисленное значение m должно обеспечивать выполнение выбранной стратегии.

Наиболее комплексным, с точки зрения эффективности функционирования системы массового обслуживания, является правило перехода, обеспечивающее минимум общих потерь от пребывания требований в системе массового обслуживания и от простоев каналов этой системы.

Условимся, что для системы, которая содержит универсальные и специализированные каналы, средняя стоимость потерь от пребывания в системе одного требования в единицу времени равна C. Тогда за средний срок, в течение которого одно требование первого типа, поступающее на специализированный канал, находится в очереди, потери от его ожидания в среднем составят Cl_1 / λ_1.

Средняя стоимость потерь из-за ожидания в очереди требования, поступающего на универсальный канал, за единицу времени равна Cl_2 / λ_2, что с учетом формулы (5а) может быть выражено как $Cl_2(\lambda_{22} + \lambda_1 P_{n+m}\mu_{12} / \mu_{22})^{-1}$.

Для вычисления потерь от всего периода пребывания требования в системе необходимо учитывать, что обслуживание одного требования первым и вторым каналом длится в среднем соответственно μ_1^{-1} и μ_2^{-1}.

Примем расходы по содержанию специализированного канала в течение единицы времени за $Ц_1$, и, следовательно, за промежуток времени, равный единице, потери от его простоев в условиях простейшей схемы (см. рис. 1) в среднем составят $P_0 Ц_1$. Средние потери из-за простоев универсального канала в тех же условиях, очевидно, составляют $Q_0 Ц_2$, где $Ц_2$ — затраты на содержание одного такого канала в единицу времени. При использовании не простейшей, а более общей схемы (см. рис. 2), средние суммы общих потерь от простоев специализированных и универсальных каналов в единицу времени соответственно составят

$$Ц_1 \sum_{k=0}^{n} P_k (n-k) \quad \text{и} \quad Ц_2 \sum_{k=0}^{N} Q_k (N-k).$$

Найти m для правила перехода, устанавливающего наименьшие потери при заданном режиме поступления и обслуживания требований в системе, можно исходя из минимизации Φ — функции потерь

$$\Phi = C\left(\frac{l_1}{\lambda_1} + \frac{1}{\mu_1} + \frac{l_2}{\lambda_{22} + \lambda_1 P_{n+m}\dfrac{\mu_{12}}{\mu_{22}}} + \frac{1}{\mu_2}\right) K +$$

$$+ T_{\text{р}}\left[Ц_1 \sum_{k=0}^{n} P_k (n-k) + Ц_2 \sum_{k=0}^{N} Q_k (N-k)\right], \qquad (11)$$

где K — количество требований, поступающих в систему за определенный период времени; $T_{\text{р}}$ — число единиц времени в пределах указанного периода, в течение которых могут работать обслуживающие каналы.

Проиллюстрируем на примере получение m для данной стратегии.

Пример 4. Определить m, соответствующее минимуму функции потерь (11), для системы с той же схемой и с теми же параметрами поступления и обслуживания, что и в примере 2, если $C = 0{,}2$ руб.; $Ц_1 = 4$ руб.; $Ц_2 = 5$ руб.; количество требований, поступающее в систему в среднем за год, $K = 4500$; среднегодовое число рабочих дней системы $T_{\text{р}} = 260$.

Решение. Значения средних времен ожидания l_1 / λ_1 и l_2 / λ_2 получены в примере 2. По условию средняя продолжительность обслуживания одного требования равна $1/\mu_1 = 1/\mu_2 = 0{,}06$ (дня).

Поскольку рассматриваемая в примере схема — простейшая, то для определения средних простоев каналов системы необходимо знать вероятности P_0 и Q_0.

P_0, исходя из (1), составляет $(10 - 8{,}1\cdot 0{,}9^m)^{-1}$, а величина Q_0, согласно (7), есть разность $1 - \alpha_2$, значение которой получено в примере 2 и равно $(3{,}6 - 3{,}7\cdot 0{,}9^m) \times$ $\times (10 - 8{,}1\cdot 0{,}9^m)^{-1}$.

После подстановки значений входящих в нее параметров функция потерь для данного примера приобретает вид

$$\Phi = 0{,}2\left[\frac{5{,}4(1 - 0{,}1m0{,}9^m - 0{,}9^m)}{10 - 8{,}1\cdot 0{,}9^m} + \frac{(6{,}4 - 4{,}4\cdot 0{,}9^m)}{(3{,}6 - 3{,}7\cdot 0{,}9^m)(110 - 102{,}5\cdot 0{,}9^m)} + 0{,}12\right]4500 +$$

$$+ 260\,\frac{4 + 5(3{,}6 - 3{,}7\cdot 09^m)}{10 - 8{,}1\cdot 0{,}9^m}.$$

Построив ряд последовательного приближения, находим, что минимальному значению Φ соответствует $m = 4$.

Таковы три стратегии, которые могут применяться в целях повышения эффективности функционирования системы массового обслуживания при загрузке требованиями ее каналов, имеющих различную дисциплину обслуживания.

ЛИТЕРАТУРА

1. Д. Р. Кокс, У. Л. Смит, Теория очередей. М., «Мир», 1966.
2. E. Koenigsberg. On jockeying on queues. Management Science, 1966, v. 12, № 5.
3. Е. С. Вентцель. Теория вероятностей. М., «Наука», 1964.
4. Б. В. Гнеденко, И. Н. Коваленко. Введение в теорию массового обслуживания. М., «Наука», 1966.
5. А. Кофман, Р. Крюон. Массовое обслуживание. Теория и приложения. М., «Мир», 1965.
6. А. А. Шахбазов, Э. Г. Самандаров. Об обслуживании неординарного потока. В сб. Кибернетику — на службу коммунизму, т. II, М., «Энергия», 1964.

Поступила в редакцию
30 VI 1967

О ЦЕЛЕСООБРАЗНОСТИ ПРОДОЛЖЕНИЯ СЛУЧАЙНОГО ПОИСКА ГЛОБАЛЬНОГО ЭКСТРЕМУМА МНОГОЭКСТРЕМАЛЬНОЙ ФУНКЦИИ

Э. М. ВАЙСБОРД

(Москва)

В статье рассматривается задача определения глобального экстремума многоэкстремальной функции. При этом применяется метод случайного выбора начальной точки с последующим градиентным спуском из этой точки; даются оценки для среднего числа итераций, необходимых для улучшения результата, достигнутого после уже проведенных итераций. Полученные оценки позволяют в некоторых случаях делать заключения о целесообразности дальнейших итераций.

1. ПОСТАНОВКА ЗАДАЧИ

1. Пусть в компактной области S m-мерного векторного пространства E_m задана непрерывно дифференцируемая функция $f(x) \equiv f(x_1,\ldots,x_m)$. Мы хотим отыскать глобальный экстремум (для определенности пусть это будет минимум) функции $f(x)$ в области S.

Если функция $f(x)$ и область S выпуклы, то при решении этой задачи очень хорошо действует градиентный метод. Однако в общем случае однократное применение градиентного метода приводит лишь к локальному минимуму функции $f(x)$. Поэтому для отыскания глобального минимума многоэкстремальной функции $f(x)$ применяют градиентный спуск несколько раз, начиная всякий раз из новой точки. Так как ни расположение, ни размеры области притяжения глобального минимума функции обычно не известны, то выбор начальной точки всякий раз естественно производить случайным образом в соответствии с равномерным распределением вероятностей на области S. После окончания каждой итерации нужно сравнить результат, достигнутый в этой итерации, с наилучшим результатом, достигнутым в прошлых итерациях. Если вновь полученный результат окажется лучше, то он оставляется в качестве результата для дальнейшего сравнения. Если же он окажется хуже, то в качестве результата для дальнейшего сравнения оставляется прежний результат. После проведения нескольких итераций имеет смысл оценить целесообразность проведения дальнейших итераций. Целесообразность при этом понимается в следующем смысле: какое в среднем число итераций необходимо еще провести, чтобы добиться улучшения результата по сравнению с уже достигнутым.

Точная постановка задачи такова. Пусть проведено n итераций градиентного спуска. Минимальное значение функции $f(x)$, полученное после этих итераций, равно f_n.

В одной из последующих итераций мы сможем получить значение функции $f(x)$ меньше, чем f_n, если начальная точка этой итерации попадает в область притяжения такой стационарной точки функции $f(x)$, для которой значение функции $f(x)$ меньше, чем f_n. Вероятность попадания начальной точки в одну из их областей притяжения