

УДК 621.371

А. А. Г р е ш и л о в, Б. П. Н а з а р е н к о,
П. А. П л о х у т а

О ПЕЛЕНГАЦИИ ИСТОЧНИКОВ ИЗЛУЧЕНИЙ

Рассмотрена некорректная задача оценки пеленгов нескольких излучателей. Проанализированы возможные методы решения (методы регуляризации) и сформулирован алгоритм решения этой задачи, который позволяет учесть погрешности всех полученных в результате измерений величин (конфлюэнтный анализ) и найти не только точечные, но и интервальные оценки углов пеленга.

Рассмотрим задачу пеленгации источников излучения посредством обработки данных, поступающих с выходов элементов антенной системы (или детекторов излучения). Эта задача находит применение во многих практических приложениях, таких как гидролокация, радиолокация, сейсморазведка и радиоастрономия. Информация, которую обычно стремятся получить о волновом поле, включает в себя следующее: число излучателей энергии и их пеленги (или пространственный энергетический спектр), сигналы, сгенерированные этими излучателями, и изменение во времени всех перечисленных характеристик. Причем для каждого параметра необходимо получить не только точечные оценки, но и ковариационную матрицу оценок или, по крайней мере, дисперсии оценок. Использование набора детекторов вместо одного предоставляет возможность улучшения соотношения сигнал-шум, подавления искажений и другие.

В настоящей статье сформулирована задача пеленгации излучателей, дан обзор наиболее известных алгоритмов ее решения и проанализированы их некоторые ограничения. В примерах рассмотрены задачи пеленгации нескольких гармонических сигналов с постоянной во времени амплитудой, различающихся амплитудой и углом прихода, с помощью регулярной линейной антенной решетки с M элементами.

Математическая модель задачи соответствует модели из работ [1,2]. Пусть набор детекторов содержит M ненаправленных детекторов, расположенных на одинаковом расстоянии d друг от друга вдоль оси x (рис. 1). Фазовый центр набора помещен в начало координат, тогда координата m -го детектора $p_m = (m - (M + 1) / 2) d$, $m \in \{1, \dots, M\}$.

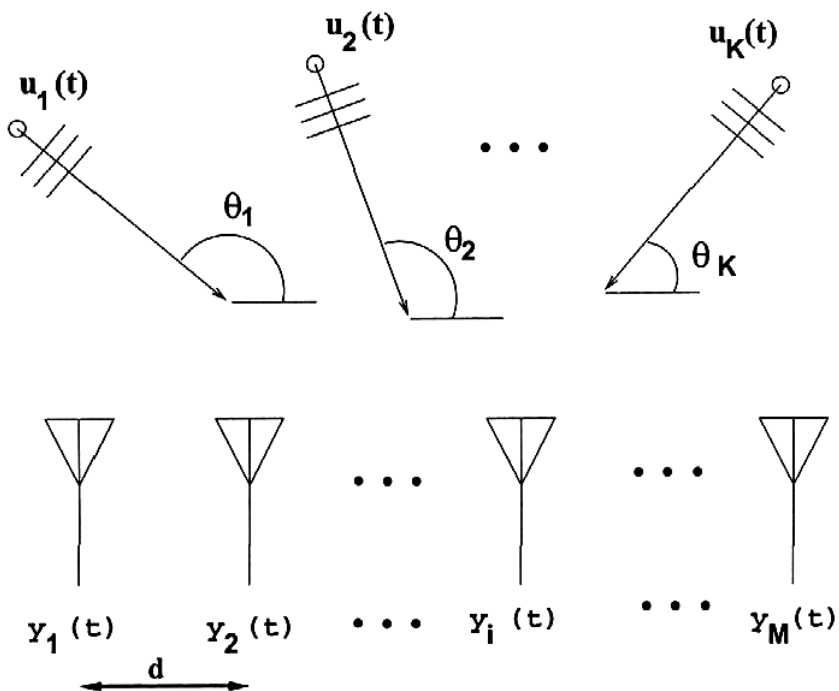


Рис. 1. Иллюстрация геометрии пеленгации излучателей:

излучатели $u_k(t)$, находящиеся по отношению к набору детекторов под углами θ_k , вызывают выходные сигналы элементов приемной антенны $y_m(t)$

Предположим, что неизвестное число пеленгуемых излучателей находится на большом расстоянии от набора детекторов. Для этих излучателей кривизна фронта волны незначительна по сравнению с расстоянием между детекторами и допустима аппроксимация волны плоским фронтом. Решение волнового уравнения для одного источника, излучающего сигнал $f(t)$, имеет вид $f(t - \mathbf{p}^T \boldsymbol{\alpha})$, где \mathbf{p} — вектор положения и $\boldsymbol{\alpha}$ — так называемый вектор задержки, ориентированный в соответствии с направлением распространения волны; его модуль равен $1/c$, т.е. величине, обратной скорости распространения сигнала. Ослабление сигнала с расстоянием не учитывается, так как оно постоянно для всего набора детекторов.

Принимаемый узкополосный сигнал может быть представлен выражением $u(t) \exp(j\omega_0 t)$, где $u(t)$ — модулирующий сигнал; ω_0 — несущая частота, которая должна быть намного выше, чем полоса пропускания $u(t)$ для регистрируемого узкополосного сигнала. Для того чтобы избежать пространственного искажения, расстояние между детекторами не должно превышать половины длины волны, т.е. $d \leq \lambda/2 = 2\pi c/(2\omega_0)$. Если не оговорено иное, выбирается $d = \lambda/2$. Выход m -го детектора равен $y_m(t) = u(t - \tau_{center}) \exp(j(\omega_0(t - \tau_{center}) - \mathbf{k}^T \mathbf{p}))$, где τ_{center} — задержка прохождения сигнала между излучателем и фазовым центром, волновое число задано вели-

чиной $\mathbf{k} = \omega_0 \boldsymbol{\alpha}$. Предположение о том, что сигнал узкополосный, позволяет не учитывать задержку $\mathbf{k}^T \mathbf{p}$ в показателе экспоненты распространения сигнала $u(t)$ между детекторами; она имеет место только при модуляции. Комплексная огибающая выхода детектора m (т.е. выход после демодуляции) может быть записана так: $y_m(t) = u(t - \tau_{center}) \exp(-j(\omega_0 \tau_{center} + \mathbf{k}^T \mathbf{p}))$. Если измерять время относительно момента поступления сигнала в фазовый центр, зависимость от τ_{center} можно не учитывать. Таким образом, для единичного излучателя, комплексная огибающая выходов детекторов может быть записана как $\mathbf{y}(t) = \mathbf{a}(\theta) u(t)$. Вектор $\mathbf{a}(\theta) = \exp(-j\mathbf{k}^T \mathbf{p})$ включает в себя информацию о фазовой задержке для излучателя, пеленг которого равен θ . Параметризация вектора $\mathbf{a}(\theta)$ по θ может быть выполнена, так как $\mathbf{k}^T \mathbf{p}_m = -(\omega_0/c)(m - (M + 1)/2) d \cos \theta$.

Поскольку система линейна, применим принцип суперпозиции, тогда модель для K узкополосных сигналов с одинаковой несущей частотой может быть записана как $\mathbf{y}(t) = \mathbf{A}(\theta) \mathbf{u}(t)$. Матрица $\mathbf{A}(\theta)$ размерности $M \times K$ состоит из векторов $\mathbf{a}(\theta)$ для различных излучателей, входящих в нее в качестве столбцов: $\mathbf{A}(\theta) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \dots, \mathbf{a}(\theta_K)]$. Вектор сигнала детектора $\mathbf{y}(t)$ — это вектор-столбец, m -м элементом которого является $y_m(t)$, и аналогично $\mathbf{u}(t)$ — это вектор-столбец, состоящий из сигналов $u_k(t)$, получаемых от всех K излучателей. Вектор $\boldsymbol{\theta}$ состоит из пеленгов всех K излучателей: $\boldsymbol{\theta} = [\theta_1, \dots, \theta_K]^T$. Принимая во внимание неизбежное присутствие помех и дискретизуя сигналы, получаем окончательный вариант модели:

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta}) \mathbf{u}(t) + \mathbf{n}(t), \quad t \in \{1, \dots, L\}. \quad (1)$$

Для простоты предполагаем, что помехи являются постоянным в пространстве и времени белым шумом, некоррелирующим с излучателями и являющимся кругообразно симметричным. Ковариационная матрица принимает следующий вид: математическое ожидание $E[\mathbf{n}(t_1) \mathbf{n}^T(t_2)] = \sigma^2 \mathbf{I} \delta(t_1 - t_2)$, где $\delta()$ — дельта-функция, \mathbf{I} — единичная матрица. Кругообразная симметрия помех приводит к тому, что $E[\mathbf{n}(t_1) \mathbf{n}^T(t_2)] = 0$.

В случае, когда излучатели находятся на небольшом расстоянии от набора детекторов, решение сферического волнового уравнения на расстоянии r от единичного источника $f(t)$ имеет вид $f(r, t) = (1/r) f(t - r/c)$. Кроме того, поскольку узкополосный сигнал может быть записан как $f(t) = u(t) \exp(j\omega_0 t)$, комплексная огибающая выхода набора детекторов принимает вид $(1/r_m) u(t - r_m/c) \times \exp(-j(\omega_0 r_m/c))$. Здесь r_m — расстояние от излучателя до m -го детектора. Пусть r_c — расстояние от излучателя до фазового центра массива, а \mathbf{p}_m — местоположение m -го детектора. Принимая

во внимание предположение об узкополосности сигналов и выбор начала отсчета времени, соответствующего прибытию сигнала в фазовый центр, выход m -го детектора можно записать как $y_m(t) = (1/r_m)u(t) \exp(-j(\omega_0(r_m - r_c)/c = a(\mathbf{p}_m)u(t)$. Здесь m -й компонент вектора $a(\mathbf{p}_m)$ содержит фазу и коэффициенты затухания для излучателя, сигнал от которого приходит на детектор m . С учетом принципа суперпозиции модель для K излучателей эквивалентна (1), за исключением того, что столбцы матрицы $\mathbf{A}(\theta)$ содержат векторы $\mathbf{a}(\theta)$ для излучателей, находящихся на небольшом расстоянии от набора детекторов, вместо векторов $\mathbf{a}(\theta)$ для сильно удаленных излучателей. В случае, когда излучатели находятся на небольшом расстоянии, отклик набора детекторов зависит не только от направлений на излучатели, но и от расстояний до них. Таким образом, в пеленгации участвуют оба эти параметра.

В случае широкополосных сигналов модель для узкополосных сигналов с помощью преобразования Фурье может быть применена для каждой частоты:

$$\mathbf{y}(\omega) = \mathbf{A}(\theta, \omega) \mathbf{u}(\omega) + \mathbf{n}(\omega), \quad \omega \in \{\omega_1, \dots, \omega_W\},$$

где $\mathbf{y}(\omega)$ и $\mathbf{u}(\omega)$ — преобразования Фурье от $\mathbf{y}(t)$ и $\mathbf{u}(t)$ соответственно. Следует обратить внимание на то, что для узкополосных сигналов имеет место только одна матрица $\mathbf{A}(\theta)$, тогда как в случае широкополосных сигналов каждой частотной составляющей ω соответствует своя матрица $\mathbf{A}(\theta, \omega)$.

Существует достаточно много алгоритмов пеленгации излучателей. Классический алгоритм основан на сканировании мощности сигнала, поступающего из различных точек, посредством управления набором детекторов, которое осуществляется при помощи компенсации задержек для выходов различных детекторов посредством соответствующего сдвига сигналов. Когда веса всех детекторов единичны и не вносятся задержки, набор детекторов работает в направлении, перпендикулярном его оси. Для волн, распространяющихся в указанном направлении, задержки относительно всех детекторов одинаковы, задержки относительно фазового центра равны нулю и не требуют компенсации. Таким образом, единичная взвешенность создает конструктивную интерференцию выходов детекторов, при этом в указанном направлении достигается максимум мощности.

Аналогично, если набор детекторов установлен под углом θ , сигнал на m -м детекторе форсирован или задержан на время $-\tau_m(\theta)$ (отрицательная задержка относительно фазового центра). Максимум мощности достигается ориентированием набора детекторов в направлении, вдоль которого происходит распространение сигнала, при усло-

вии отсутствия отражений. Задержки равны фазовым сдвигам, которые могут быть представлены комплексными весами \mathbf{w} детекторов. Таким образом, выход набора детекторов можно записать как $z(t) = \mathbf{w}^T \mathbf{y}(t) = \mathbf{w}^T \mathbf{a}(\theta) u(t)$. Для того чтобы “повернуть” набор детекторов на угол θ , веса должны быть установлены так: $\mathbf{w} = \mathbf{a}(\theta)$. Поскольку система линейна, аналогичный подход может быть использован для нахождения суперпозиции плоских волн, приходящих с разных направлений и имеющих одинаковые несущие частоты. Классический спектр может быть представлен в виде

$$P_{bf}(\theta) = \sum_{t=1}^L \|\mathbf{w}^T(\theta) \mathbf{y}(t)\|_2^2. \quad (2)$$

Классический алгоритм — это очень простой и робастный подход, который широко применяется на практике. Однако он имеет недостаток, связанный с ограничением релейского разрешения [1], который может быть смягчен только увеличением числа детекторов; улучшение соотношения сигнал-шум или увеличение времени наблюдения не изменяют разрешения.

Кроме классического алгоритма, применяются Capon’s (MVDR) [3], MUSIC [4], алгоритмы, основанные на методе максимума правдоподобия (ММП) [2, 5] и другие. Наиболее приемлемы алгоритмы, основанные на методе максимума правдоподобия. Они принадлежат к классу параметрических алгоритмов, где оцениваются параметры падающей волны. Помеха моделируется как стационарный гауссовский сигнал. Для ММП целью является поиск θ , $\mathbf{u}(t)$ и σ^2 , доставляющих максимум функции правдоподобия:

$$L_{DML}(\theta, \mathbf{u}(t), \sigma^2) = \prod_{t=1}^T (\pi\sigma^2)^{-M} \exp\left(-\|\mathbf{y}(t) - \mathbf{A}(\theta)\mathbf{u}(t)\|_2^2 / \sigma^2\right), \quad (3)$$

где θ — вектор пеленгов излучателей. Логарифмическая функция правдоподобия имеет вид

$$l_{DML}(\theta, \mathbf{u}(t), \sigma^2) = -2M \log \sigma + \frac{1}{\sigma^2 T} \sum_{t=1}^T \left(-\|\mathbf{y}(t) - \mathbf{A}(\theta)\mathbf{u}(t)\|_2^2\right). \quad (4)$$

Вычислительная сложность данного алгоритма значительно превышает сложность классического алгоритма. Преимущество алгоритмов ММП-семейства состоит в возможности различать когерентные сигналы и в обеспечении более хороших статистических свойств [6]. Кроме того, с помощью теоремы Крамера–Рао ММП позволяет получить ковариационную матрицу оценок, что в ряде случаев бывает принципиально важным [12, 14].

Задача оценки пеленга — типичная некорректная обратная задача. Теоретически сумма гармонических сигналов на одной частоте дает гармонический сигнал. Цель в задачах пеленгации — найти “слагаемые” по зарегистрированной “сумме”.

Рассмотрим линейные некорректные задачи. Позже задача пеленгации излучателя будет сведена к задаче данного типа. В обратных задачах [7–9] функция от неизвестной величины, которую необходимо определить, известна. Цель — найти имеющую физический смысл обратную функцию. Имеем $y = T(x)$, где вектор $x \in \chi$ известен, а $y \in Y$ — вектор измерений. Обычно T — это удобный для анализа непрерывный оператор, и решение прямой задачи (нахождение y при заданном x) не сталкивается с серьезными трудностями. Обратное отображение от y к x в интересующих нас задачах гораздо сложнее, что связано с недоопределенностью задачи, неоднозначностью решений или разрывной зависимостью решения от результатов измерений. Наличие любой из перечисленных проблем делает задачу некорректной.

В постановке, изложенной выше, задача является слишком общей. Примем дополнительные допущения. Пусть χ и Y — конечномерные пространства, а T — линейный оператор:

$$y = Tx, \quad y \in \mathbb{C}^M, \quad x \in \mathbb{C}^N, \quad T \in \mathbb{C}^{M \times N}. \quad (5)$$

Положим, что задача (5) — некорректная. Классический способ ее решения — вычисление псевдообратного оператора T^\dagger . Применяя псевдообратный оператор, находим решение с минимальным квадратом нормы. Поскольку T^\dagger — линейная функция в конечномерном пространстве, то оператор обязательно непрерывен. Однако в некоторых практических задачах число обусловленности N оператора T^\dagger может быть очень большим, что делает псевдообратный оператор неприменимым на практике, так как в этом случае добавление даже небольшой помехи к измерениям может привести к физически неприемлемому решению.

Эффективное решение некорректных обратных задач основано на методах регуляризации. Регуляризация используется для решения некорректных задач посредством объединения априорных знаний об x для стабилизации задачи и предоставления корректных и практически пригодных решений. Метод состоит в минимизации некоторой меры $J_1(x)$ близости y и множества значений Tx и, в то же время, удовлетворении, насколько это возможно, условия близости x и априорной информации о нем (посредством минимизации некоторой меры $J_2(x)$). Таким образом, задача сводится к многокритериальной задаче математического программирования. Две целевые функции не

могут быть минимизированы одновременно. Необходимо найти компромисс, который мог бы быть легко достигнут при помощи линейной комбинации двух функций

$$J(\mathbf{x}) = J_1(\mathbf{x}) + \lambda J_2(\mathbf{x}). \quad (6)$$

Скаляр λ — это параметр регуляризации, регулирующий соотношение между $J_1(\mathbf{x})$ и $J_2(\mathbf{x})$. Если $\lambda = 0$, то получается нерегуляризованное решение. Если λ велико, решение определяется априорной информацией об \mathbf{x} . В общем случае выбор подходящего λ связан с особенностями решаемой задачи и является нетривиальным.

Один из наиболее широко известных подходов к регуляризации предложен Тихоновым [10]. Алгоритм Тихонова предполагает, что норма решения должна быть малой, это ограничивает коэффициент усиления в соответствии с малыми собственными числами. Целевая функция принимает вид

$$J(\mathbf{x}) = \|\mathbf{T}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_2^2. \quad (7)$$

Здесь норма разности ℓ_2 есть условие близости к данным измерений $J_1(\mathbf{x})$, а $\|\mathbf{x}\|_2^2$ играет роль $J_2(\mathbf{x})$ из соотношения (6). Значение оценки $\hat{\mathbf{x}}$, доставляющее минимум целевой функции (7), может быть вычислено по следующей формуле [10]:

$$\hat{\mathbf{x}} = \sum_{i=1}^K \left(\frac{\sigma_i^2}{\sigma_i^2 + \lambda} \right) \frac{\mathbf{u}'_i \mathbf{y}}{\sigma_i} \mathbf{v}_i. \quad (8)$$

Другие алгоритмы квадратичной регуляризации основаны на соотношении из работы [11] $\hat{\mathbf{x}} = \sum_{i=1}^K w_i \frac{\mathbf{u}'_i \mathbf{y}}{\sigma_i} \mathbf{v}_i$. Они могут рассматриваться как взвешенные псевдообращения с весами w_i . Регуляризация Тихонова — это частный случай при $w_i = \left(\frac{\sigma_i^2}{\sigma_i^2 + \lambda} \right)$. Идея всех указанных алгоритмов заключается в том, что большинство больших собственных чисел оставляют неизменными и ограничивают влияние маленьких собственных чисел матрицы \mathbf{T} , поскольку помеха усиливается в \sqrt{N} раз в направлении собственного вектора, соответствующего наименьшему собственному числу (здесь N — отношение наибольшего собственного числа матрицы \mathbf{T} к наименьшему). Алгоритмы квадратичной регуляризации очень удобны во многих практических задачах, и их преимущество состоит в том, что решение имеет короткую аналитическую запись. Однако, так как обратный оператор всегда является линейной функцией измерений, имеют место некоторые ограничения. В частности, в связи с линейностью невозможно восстановить

часть \mathbf{x} , которая принадлежит нуль-пространству \mathbf{T} (острые всплески сигнала \mathbf{x}). Данный факт показывает необходимость разработки более общего нелинейного вида регуляризации, который сможет обеспечить серьезное улучшение характеристик в данном отношении, сохраняя острые пики и другие важные элементы, а также обеспечить сверхразрешение при восстановлении решения [11]. Вычислительная сложность в последнем случае значительно возрастает.

Две известные не квадратичные регуляризирующие функции — это полная вариация и энтропия [9]. Полная вариация накладывает штраф на сумму вариаций сигнала $J_2 = \|\Delta\mathbf{x}\|_1 = \sum_i |[\Delta\mathbf{x}]_i|$, где Δ — дискретная аппроксимация оператора градиента. Полная вариация наиболее часто используется в задачах обработки изображений, в частности для их восстановления. По сравнению с регуляризацией Тихонова штраф на мощные составляющие меньше, и восстановленное решение может содержать острые пики.

Регуляризация посредством максимальной энтропии использует следующую функцию, подобную энтропии: $J_2(\mathbf{x}) = \sum_i |x_i| \log |x_i|$.

Возможны некоторые вариации с использованием перекрестной энтропии и дивергенции. Регуляризирующие функции в данной форме приводят к большей концентрации энергии в восстановленном решении (большинство коэффициентов очень малы, несколько — велики). Другой способ концентрации энергии — это некорректность. Данный способ лучше всего применим для данных, которые обладают похожим характером изменения, например, спектральная оценка для сигналов с несколькими гармониками или пеленгация точечного излучателя.

Другая регуляризирующая функция, которая обладает таким же эффектом некорректности, — это ℓ_1 -штраф: $J_2 = \|\mathbf{x}\|_1$. Данная функция подобна полной вариации, за исключением того, что берется ℓ_1 -норма значений \mathbf{x} вместо их производных. Полная вариация допускает некорректные скачки градиента \mathbf{x} , тогда как ℓ_1 -штраф выделяет некорректные значения \mathbf{x} . Более того, для снижения штрафа, накладываемого на большие по значению составляющие, несколько невыпуклых функций также нашли применение, например, ℓ_p -квазинорма с $p < 1$. $J_2 = \|\mathbf{x}\|_p^p = \sum_i |x_i|^p$ обеспечивает более сильную концентрацию энергии в восстановленном решении. Свойство сильной концентрации энергии делает ℓ_1 - и ℓ_p -регуляризации очень удобными для восстановления некорректных сигналов. Существуют и другие методы регуляризации [12].

Выбор подходящего алгоритма регуляризации зависит от конкретной задачи. Во многих математических обратных задачах выбор делается на основе ограничений различных видов, связанных с гладкостью

или энергией, а в качестве соответствующих регуляризирующих функций используются l_2 -нормы x или его производных.

Априорная некорректность удобна при поиске некорректных сигналов x . Некорректность вектора x определяется как наличие в нем небольшого числа больших по значению элементов и равенство нулю остальных элементов. Важной линейной обратной задачей, для решения которой хорошо подходит априорная некорректность, является задача представления сигнала при помощи переопределенных базисов. Их основой являются работы Маллата [15], Донохо [16], Пао [17] и других, касающиеся аппроксимации функций и оптимального выбора базиса.

В течение последнего десятилетия задаче выбора подходящего базиса для семейства сигналов было уделено много внимания и введено множество новых базисов, таких как wavelets-базис, ridgelets-базис, curvelets-базис и многие другие. Несмотря на то, что посредством любого базиса, определенного на интервале в конечномерном пространстве, можно хорошо представить любой сигнал из данного пространства, тем не менее для конкретных задач некоторые базисы обладают лучшими аппроксимационными свойствами, чем какие-либо другие.

Практические задачи, в которых выгодно некорректное представление, — это задачи сжатия, фильтрации помех и параметрического оценивания. В задачах сжатия для передачи информации (если представление сигнала не является некорректным) необходимо передавать весь сигнал. Однако, если при помощи замены базиса представление стало некорректным, возможна реальная экономия. Большинство коэффициентов разложения очень мало (по определению некорректности) и, если им присваиваются нулевые значения, качество восприятия сигнала уменьшается незначительно. Таким образом, для передачи остаются только большие по значению коэффициенты, число которых мало. Эта идея нашла применение в коммерческих алгоритмах сжатия. Другая задача, в которой некорректность играет ключевую роль, — это фильтрация помех. Если сигнал некорректен, то отделение его от помех требует значительно меньших усилий, по сравнению с ситуацией, когда мощность сигнала равномерно распределена по всей полосе сигнала. Следовательно, чтобы облегчить фильтрацию помех для класса сигналов, необходимо найти базис, в котором представление всех сигналов, принадлежащих к данному классу, настолько некорректно, насколько это возможно.

Число некорректно представимых сигналов непосредственно зависит от числа элементов в базисе. Все минимальные базисы имеют одинаковое число некорректно представимых сигналов. В целях увеличения числа сигналов, имеющих некорректное представление, используются переопределенные базисы. Некоторые переопределенные

базисы включают в себя конкатенацию нескольких ортогональных базисов, например стандартный базис и базис Фурье, которые могут некорректно представлять суперпозиции непрерывных синусоид и явление локальных всплесков. Другая возможность — это использование одного расширенного неортогонального базиса, такого как базис Фурье с числом рассматриваемых частот, превышающим размерность пространства. Переопределенный базис Фурье позволяет некорректно представить гармоники с частотами, лежащими между стандартными частотами Фурье.

Для построения переопределенного базиса следует оценить число источников в зарегистрированном сигнале по алгоритмам, описанным в работах [2, 12, 13].

При использовании переопределенных базисов теряется очень важное свойство — уникальность представления. Для восстановления уникальности ищется наиболее некорректное решение среди множества возможных решений. Математически при отсутствии шума задача формулируется так: дан сигнал $\mathbf{y} \in \mathbb{C}^M$ и переопределенный базис $\mathbf{T} \in \mathbb{C}^{M \times N}$; необходимо найти $\mathbf{x} \in \mathbb{C}^N$, такой, что $\mathbf{y} = \mathbf{T}\mathbf{x}$ и вектор \mathbf{x} — некорректен. Определим $\|\mathbf{x}\|_0^0$ как число ненулевых элементов \mathbf{x} . Необходимо найти минимум $\|\mathbf{x}\|_0^0$ при условии $\mathbf{y} = \mathbf{T}\mathbf{x}$, что является сложной комбинаторной задачей. Можно показать, что при некоторых условиях, накладываемых на \mathbf{T} и \mathbf{x} , оптимальное решение данной проблемы может быть найдено точно при помощи решения связанной задачи: найти $\min \|\mathbf{x}\|_p^p$ при условии $\mathbf{y} = \mathbf{T}\mathbf{x}$, где $0 < p \leq 1$ (отдельно рассматриваются случаи $p = 1$ и общий случай $0 < p \leq 1$).

Естественное расширение для практического применения — добавление белого гауссовского шума \mathbf{n} , т.е.

$$\mathbf{y} = \mathbf{T}\mathbf{x} + \mathbf{n}. \quad (9)$$

Уравнение (9) может быть решено следующим образом [18, 19]:

$$\min \|\mathbf{y} - \mathbf{T}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_p^p. \quad (10)$$

Априорная часть $J_2(\mathbf{x})$ оказывает эффект принудительной регуляризации. На рис. 2 показано, почему ℓ_p -регуляризация при $p \leq 1$ лучше обеспечивает некорректное представление \mathbf{x} . На графике (б) — значения ℓ_p -нормы для p -й степени ($\|\mathbf{x}\|_p^p$) для $p = 0,5$, $p = 1$ и $p = 2$ двумерного вектора. По сравнению с фиксированной ℓ_2 -нормой, т.е. всеми векторами, которые лежат на окружности с фиксированным радиусом, ℓ_p -нормы при $p \leq 1$ являются минимизированными на координатных осях, т.е. соответствуют тому, что некоторые коэффициенты равны нулю, в то время как другие имеют большие значения. Данное утверждение может быть обобщено на векторы больших размерностей. График (а) показывает ℓ_p -нормы для тех же значений p

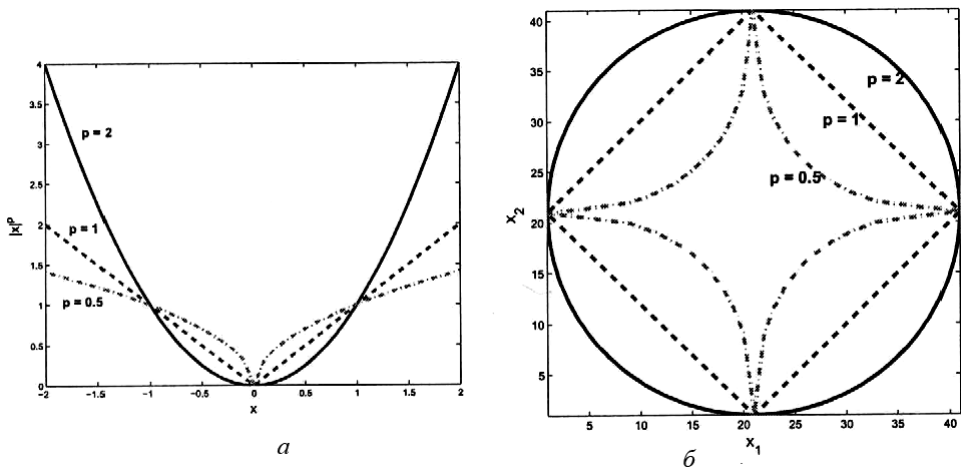


Рис. 2. Одномерный график $\|x\|_p^p$ при $p = 50, 5; 1; 2$ (а) и двумерное множество уровней $\|x\|_p^p$ при тех же p (б)

в одномерном случае, т.е. то, что штраф на большие составляющие (большие x_i) уменьшается с уменьшением p . Мощные составляющие штрафуются намного меньше при ℓ_p -штрафовании с $p \leq 1$, чем при ℓ_2 -штрафовании (регуляризация Тихонова). Этим объясняется сглаживающий эффект ℓ_2 -штрафования.

Другой вывод, который можно сделать из рис.2 — это то, что ℓ_1 -норма — выпуклая, тогда как при $p \leq 1$ ℓ_p -норма уже не выпукла. При $p \leq 1$ неравенство треугольника не выполняется, и более справедливо применить термин “квазинорма” вместо термина “норма”. Однако термин ℓ_p -норма будет использован для любого значения p . Вычислительная процедура минимизации нескольких невыпуклых регуляризирующих функций (в частности ℓ_p) может быть упрощена при использовании полуквадратичного алгоритма регуляризации [17]. Ключевая идея состоит в введении добавочного вектора s и расширенной функции стоимости $Q(x, s)$, которая является квадратичной по x при фиксированном s и минимум которой $\min_s Q(x, s) = J(x)$ для любого x . Если $Q(x, s)$ легко минимизируется по s (или если существует точное решение), то расширенная функция $Q(x, s)$ может быть оптимизирована с приемлемой эффективностью при помощи итеративного координатного спуска.

Задача представления сигнала посредством переопределенного базиса отличается от более тривиальных обратных задач, таких как нахождение оригинала изображения. Основное внимание в последних концентрируется на том, чтобы сделать обратную функцию непрерывной, что соответствует минимизации ℓ_2 -нормы x , которая не приводит к некорректным решениям. Используется другой аспект регуляризации, определяющий однозначное решение среди большого множества

возможных решений. Регуляризация — очень гибкая процедура, позволяющая использовать соответствующую априорную информацию и ℓ_p -норму при $p \leq 1$, которая обеспечивает некорректность.

Другая причина использования регуляризации — возможность смягчить допущение переопределенности. Это означает, что матрица \mathbf{T} не должна быть переопределенной и она даже не должна охватывать пространство y . Предположим, y лежит в области значений \mathbf{T} , но некоторые коэффициенты \mathbf{x} , полученные при помощи обычного обращения, очень малы. При использовании некорректной регуляризации получается семейство решений, которое позволяет найти более некорректное (но менее точное) представление y .

Кроме представления сигнала при помощи переопределенного базиса некорректная регуляризация имеет множество приложений во многих областях, таких как статистика, анализ данных и обучение машин. Важной задачей во всех трех областях является выбор подмножества. Предположим, наблюдаемая величина y зависит от множества параметров $\mathbf{x} = [x_1, \dots, x_N]$, но влияние малого подмножества параметров намного больше, чем влияние всех остальных. Для того чтобы надлежащим образом построить модель для y посредством элементов \mathbf{x} , необходимо найти малое подмножество $\{x_i\}$, которое хорошо прогнозирует y .

Сначала рассмотрим вариант задачи представления сигнала посредством переопределенного базиса без помех и опишем ее решение с использованием ℓ_1 -штрафования. Хотя вариант без помех имеет малое практическое значение, он тесно связан с формулировкой задачи с учетом помех, и большинство теоретических результатов получено именно для такого варианта.

Предположим, что существует сигнал y , который является линейной комбинацией нескольких элементов переопределенного базиса \mathbf{A} :

$$y = \sum_{i=1}^K x_i \mathbf{a}_i, \text{ где } \mathbf{x} = \mathbf{x}^+ - \mathbf{x}^-, \mathbf{x}^+ \geq 0, \mathbf{x}^- \geq 0, \mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_N].$$

Найдем весовые коэффициенты x_i , используя доступную информацию: y и \mathbf{A} . Так как \mathbf{A} — переопределенный базис, то нахождение x_i — это некорректная обратная задача (существует огромное множество решений). Логично найти решение, решая следующую задачу:

$$\min \|\mathbf{x}\|_0 \text{ при ограничении } y = \mathbf{A}\mathbf{x}. \quad (11)$$

Поскольку $\|\mathbf{x}\|_0$ представляет число ненулевых элементов \mathbf{x} , эта сложная комбинаторная задача заменяется задачей

$$\min \|\mathbf{x}\|_1 \text{ при ограничении } y = \mathbf{A}\mathbf{x}. \quad (12)$$

Выражение (12) может быть переформулировано как задача линейного программирования, в случае если данные вещественные, и как задача математического программирования второго порядка, если данные комплексные.

Сначала рассмотрим численное решение для случая, когда данные вещественные. Процедура переформулирования задачи приводит к задаче минимизации ℓ_1 -нормы, которая является широко известной задачей линейного программирования. Введем две переменные \mathbf{x}^+ и \mathbf{x}^- , определенные как $x_i^+ = \max\{x_i, 0\}$, $x_i^- = \max\{-x_i, 0\}$, при помощи которых \mathbf{x} может быть легко восстановлен следующим образом: $\mathbf{x} = \mathbf{x}^+ - \mathbf{x}^-$. Переменные \mathbf{x}^+ и \mathbf{x}^- ограничены положительным октантом и должны удовлетворять условию $x_i^+ x_i^- = 0, \forall i$. Оно автоматически выполняется, когда рассматривается задача минимизации. Переформулировав задачу (12), получим

$$\min 1' \begin{pmatrix} \mathbf{x}^+ \\ \mathbf{x}^- \end{pmatrix} \text{ при ограничениях} \\ (\mathbf{A} \quad -\mathbf{A}) \begin{pmatrix} \mathbf{x}^+ \\ \mathbf{x}^- \end{pmatrix} = \mathbf{y} \text{ и } \begin{pmatrix} \mathbf{x}^+ \\ \mathbf{x}^- \end{pmatrix} \geq 0. \quad (13)$$

Численное решение задачи (13) легко найти при помощи симплекс-метода или метода внутренних точек при линейном программировании.

Задачу (12) без учета помех можно представить в следующем виде:

$$\min \|\mathbf{x}\|_1 \text{ при ограничении } \|\mathbf{y} - \mathbf{Ax}\|_2^2 = 0. \quad (14)$$

С учетом помех уравнение (14) принимает вид:

$$\min \|\mathbf{x}\|_1 \text{ при ограничении } \|\mathbf{y} - \mathbf{Ax}\|_2^2 \leq \beta^2, \quad (15)$$

где β — параметр регуляризации, который характеризует желаемую интенсивность помехи. Также может быть предложено несколько альтернативных формулировок задачи. Наиболее широко распространенная формулировка (она, в частности, использована в работах [15, 19]) — это штрафная форма

$$\min \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda \|\mathbf{x}\|_1. \quad (16)$$

Если же поменять местами целевую функцию и ограничение, то получим третью форму записи:

$$\min \|\mathbf{y} - \mathbf{Ax}\|_2^2 \text{ при ограничении } \|\mathbf{x}\|_1 \leq \delta. \quad (17)$$

Следует обратить внимание на то, что можно заменить ограничение $\|\mathbf{y} - \mathbf{Ax}\|_2^2 \leq \beta^2$ ограничением $\|\mathbf{y} - \mathbf{Ax}\|_2 \leq \beta$, что приводит к эквивалентной штрафной форме

$$\min \|\mathbf{y} - \mathbf{Ax}\|_2 + \tilde{\lambda} \|\mathbf{x}\|_1. \quad (18)$$

Отличие выражения (18) от (16) заключается в том, что первое слабое имеет первую степень, а не вторую. Эта форма записи может быть представлена более наглядно при помощи математического программирования второго порядка.

Численный метод для ℓ_p -оптимизации с учетом помех

$$\min J(\mathbf{x}) = \min \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda \|\mathbf{x}\|_p^p \quad (19)$$

был разработан в работе [20]. Метод особенно эффективен при $p < 1$. Использование ℓ_p -алгоритмов с $p < 1$ ограничено в связи с возникающими трудностями, так как ℓ_p -норма не является выпуклой для $p < 1$ (и даже ее множества уровня не являются выпуклыми). Таким образом, не гарантируется достижение глобального минимума при помощи методов локальной оптимизации. Вместо использования алгоритмов глобальной оптимизации, чаще используют алгоритмы локальной оптимизации, имея в виду, что есть хорошее начальное приближение.

Использование ℓ_p -норм с $p < 1$ для обеспечения некорректности может быть обосновано точно так же, как и использование ℓ_1 -норм. Фактически, когда $p < 1$, необходимые условия глобального экстремума ℓ_p -задачи без помех ($\min \|\mathbf{x}\|_p^p$ при ограничении $\mathbf{y} = \mathbf{Ax}$) становятся эквивалентными условиям глобального экстремума ℓ_0 -задачи ($\min \|\mathbf{x}\|_0^0$ при ограничении $\mathbf{y} = \mathbf{Ax}$), которые являются менее жесткими. Однако нахождение глобального решения ℓ_p -задачи без помех для $p < 1$ является более сложным, чем для случая, когда $p = 1$ (ℓ_p -задача без помех недостаточно выпукла).

Перейдем к формулировке задачи с помехами: $J(\mathbf{x}) = \min \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda \|\mathbf{x}\|_p^p$. Кроме невыпуклости возникает еще одна трудность при использовании ℓ_p -целевой функции, которая заключается в том, что она не дифференцируема в $\mathbf{0}$. Фактически, целевые функции не дифференцируемы для обоих случаев: $p = 1$ и $p < 1$. Но для $p = 1$ существует способ обойти эту трудность: в случае вещественных данных представлением $\mathbf{x} = \mathbf{x}^+ - \mathbf{x}^-$, где $\mathbf{x}^+ \geq 0$, $\mathbf{x}^- \geq 0$, осуществляется переход к эквивалентной задаче линейного программирования. Для $p < 1$ это невозможно, и вместо этого рассматривают дифференцируемые аппроксимации ℓ_p -целевой функции; эти аппроксимации можно также использовать для ℓ_1 -целевой функции. Дифференцируемые аппроксимации обычно содержат параметр, который отвечает за компромисс между гладкостью аппроксимации и близостью к не дифференцируемой функции, для которой производится аппроксимация. Другими словами, дифференцируемая аппроксимация — это семейство функций.

Для $p = 1$ известна следующая аппроксимация:

$$\|\mathbf{x}\|_1 \approx \sum_{i=1}^N \rho(x_i), \quad (20)$$

где

$$\rho(x) = \begin{cases} x_i^2, & \text{если } |x_i| < \gamma; \\ |x_i| - \gamma + \gamma^2, & \text{если } |x_i| \geq \gamma; \end{cases} \quad (21)$$

γ — параметр. При $\gamma \rightarrow 0$, аппроксимация равномерно стремится к ℓ_1 -норме. Однако для любого $\gamma > 0$ она дифференцируема. Данная аппроксимация может быть переформулирована в задачу квадратичного программирования [21]. Другая аппроксимация [22], которая работает при $p < 1$, также как и при $p = 1$, имеет вид

$$\|\mathbf{x}\|_p^p \approx \sum_{i=1}^N (|x_i|^2 + \varepsilon)^{p/2}. \quad (22)$$

Здесь $\varepsilon > 0$ — параметр гладкости. Данная дифференцируемая аппроксимация хорошо подходит для поставленных целей и дает модифицированную регуляризующую функцию

$$J(\mathbf{x}) \approx J_\varepsilon(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \sum_{i=1}^N (|x_i|^2 + \varepsilon)^{p/2}. \quad (23)$$

Следует отметить, что $J_\varepsilon(\mathbf{x}) \rightarrow J(\mathbf{x})$, когда $\varepsilon \rightarrow 0$. В качестве ε берем константу малой величины. Минимизация $J_\varepsilon(\mathbf{x})$ в общем случае не приводит к решению в замкнутой форме, поэтому необходимо использовать численные алгоритмы.

Для решения данной задачи оптимизации в [20] используется алгоритм полуквадратичной регуляризации [23]. Полуквадратичная регуляризация переводит задачу неквадратичной оптимизации в совокупность квадратичных задач. Для наших целей достаточно рассмотреть квазиньютоновский алгоритм [14]. Однако доказательство локальной сходимости из любой начальной точки основано на полуквадратичных решениях алгоритма.

Используется следующий итеративный алгоритм:

$$\mathbf{H}(\hat{\mathbf{x}}^{(n)}) \hat{\mathbf{x}}^{(n+1)} = 2\mathbf{A}^T \mathbf{y}, \quad (24)$$

где n — номер итерации;

$$\mathbf{H}(\mathbf{x}) \triangleq 2\mathbf{A}^T \mathbf{A} + \lambda \Lambda(\mathbf{x}); \quad (25)$$

$$\Lambda(\mathbf{x}) \triangleq \text{diag} \left\{ \frac{p}{(|x_i|^2 + \varepsilon)^{1-p/2}} \right\};$$

$\text{diag}\{\cdot\}$ — диагональная матрица, i -й диагональный элемент которой вычисляется по формуле, заключенной в скобки. Здесь важным отличием от таких методов, как гребневая регрессия [27], является зависимость \mathbf{H} от $\Lambda(\mathbf{x})$ — варьируемого штрафа.

Алгоритм может также интерпретироваться как квазиньютоновский алгоритм с аппроксимацией гессиана. Полный гессиан определяется выражением

$$\nabla_{xx} J_\varepsilon = \mathbf{H}(\mathbf{x}) + \lambda \text{diag} \left\{ \frac{p(p-2)|x_i|^2}{(|x_i|^2 + \varepsilon)^{2-p/2}} \right\}. \quad (26)$$

Для $p < 2$ второе слагаемое всегда отрицательно и может привести к бесконечности гессиана. Оставляя только первое слагаемое, мы получаем положительно определенную аппроксимацию гессиана. Кроме того, для $p = 2$ аппроксимация становится точной.

Итеративный квазиньютоновский алгоритм имеет следующий вид:

$$\hat{\mathbf{x}}^{(n+1)} = \hat{\mathbf{x}}^{(n)} - \beta \mathbf{H}(\hat{\mathbf{x}}^{(n)})^{-1} \nabla_x J_\varepsilon(\hat{\mathbf{x}}^{(n)}). \quad (27)$$

При выборе шага $\beta = 1$ сокращение слагаемых в правой части приводит данное выражение к виду (24). Градиент $J_\varepsilon(\mathbf{x})$ выражается формулой

$$\begin{aligned} \nabla_x J_\varepsilon &= 2\mathbf{A}^T(\mathbf{A}\mathbf{x} - \mathbf{y}) + \lambda \text{vec} \left\{ \frac{px_i}{(|x_i|^2 + \varepsilon)^{1-p/2}} \right\} = \\ &= \left(2\mathbf{A}^T\mathbf{A}\mathbf{x} + \lambda \text{diag} \left\{ \frac{p}{(|x_i|^2 + \varepsilon)^{1-p/2}} \right\} \right) \mathbf{x} - 2\mathbf{A}^T\mathbf{y} = \mathbf{H}(\mathbf{x})\mathbf{x} - 2\mathbf{A}^T\mathbf{y}. \end{aligned} \quad (28)$$

Отсюда получаем

$$\hat{\mathbf{x}}^{(n+1)} = \hat{\mathbf{x}}^{(n)} - \mathbf{H}(\hat{\mathbf{x}}^{(n)})^{-1} \nabla_x J_\varepsilon(\hat{\mathbf{x}}^{(n)}) = 2\mathbf{H}(\hat{\mathbf{x}}^{(n)})^{-1} \mathbf{A}^T\mathbf{y}, \quad (29)$$

что приводит к выражению (24).

Продолжаем итеративный процесс в (24) до тех пор, пока

$$\frac{\|\hat{\mathbf{x}}^{(n+1)} - \hat{\mathbf{x}}^{(n)}\|_2^2}{\|\hat{\mathbf{x}}^{(n)}\|_2^2} < \delta,$$

где $\delta > 0$ — малая константа. Сходимость алгоритма из любой начальной точки к локальному минимуму гарантирована [28, 29].

В рассматриваемой задаче пеленгации сигналов необходимо дополнительно учесть влияние погрешностей всех величин (кроме по-

мехи зарегистрированного сигнала), входящих в уравнение падающей волны, которые не фигурировали в предыдущем анализе:

- погрешность определения расстояния между детекторами;
- точность установки направлений;
- точность определения частоты ω_0 .

Поскольку за основу принят ММП, то функции плотностей вероятностей названных случайных величин следует включить в функцию правдоподобия (3) [14] и тем самым получить функционал (логарифмическую функцию правдоподобия) для оценки искомых параметров.

ММП позволяет получить и ковариационную матрицу (матрицу рассеяния) $\mathbf{D}(\hat{\theta})$ найденных оценок $\hat{\theta}$ по матрице \mathbf{M} , составленной из вторых производных логарифмической функции правдоподобия $\ln L$ по θ , вычисленных при найденных оценках $\hat{\theta}$, т.е. $\mathbf{D}(\hat{\theta}) = \mathbf{M}^{-1}$,

$$\text{где } M_{ij} = \left. \frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \right|_{\theta = \hat{\theta}}, \quad i, j = 1, 2, \dots, K.$$

Поскольку введение переопределенных базисов позволяет свести задачу пеленгации к линейной, то рассмотрим процедуру учета погрешности элементов матрицы системы для линейных систем:

$$\eta = f(\xi, \Theta) = \sum_{i=1}^n \Theta_i \xi_i.$$

В процессе наблюдений из-за влияния погрешностей измерений δ_i и ε_{ij} имеем $x_{ij} = \xi_{ij} + \delta_{ij}$; $y_i = \eta_i + \varepsilon_i$, $i = 1, 2, \dots, n$. Таким образом, оцениваются параметры линейной модели с погрешностями в матрице системы и в правой части уравнений. Будем считать, что ошибки измерений суть независимые нормально распределенные случайные величины с нулевыми средними значениями и известными дисперсиями $\sigma^2(x_{ij})$ и $\sigma^2(y_i)$.

В этом случае логарифмическая функция правдоподобия $\ln L$, для нашего описания это функция $J_1(x)$, с точностью до констант будет иметь вид

$$F = \frac{1}{2} \sum_{j=1}^m \left\{ \sum_{i=1}^n \frac{(x_{ij} - \xi_{ij})^2}{\sigma^2(x_{ij})} + \frac{(y_j - \sum_i \Theta_i \xi_{ij})^2}{\sigma^2(y_i)} \right\}, \quad (30)$$

а ограничение, определяющее принадлежность полученных оценок “истинных” значений аргументов ξ_{ij} области их возможных значений D_{ij} ($\hat{\xi}_{ij} \in D_{ij}$) можно записать следующим образом:

$$\left| \hat{x}_{ij} - \hat{\xi}_{ij} \right| \leq 3\sigma(x_{ij}). \quad (31)$$

Здесь для упрощения вида функционала (30) ошибки измерений считаются статистически независимыми.

Задача минимизации по Θ функционала (30) при фиксированных значениях ξ есть обычная задача регрессионного анализа. Исследования показали, что для обеспечения минимального времени счета и большей точности лучше всего на первом шаге решать систему линейных уравнений при $\xi_{ij} = x_{ij}$ методом Гаусса с выбором максимального элемента, в дальнейшем минимизируя по Θ функционал (30) методом сопряженных градиентов. Учитывая вид функционала (30), задачу минимизации по Θ легко свести к задаче поиска минимума квадратичной формы

$$G(\Theta) = \frac{1}{2}\Theta^T A\Theta + a^T\Theta,$$

которая решается методом сопряженных градиентов. Здесь Θ — вектор искомых оценок параметров; A — матрица квадратичной формы, элементы которой вычисляются по формуле

$$A_{rp} = \sum_{j=1}^m \frac{1}{\sigma^2(y_j)} x_{rj} x_{pj}; \quad r = 1, 2, \dots, n; \quad p = 1, 2, \dots, n,$$

x_{rj} — значение r -й переменной, полученное в j -м измерении; a^T — вектор, компоненты которого равны

$$a_r^T = - \sum_{j=1}^m \frac{y_j}{\sigma^2(y_j)} x_{rj}; \quad r = 1, 2, \dots, n.$$

Пересчет точных значений ξ на основании условий $\frac{\partial F}{\partial \xi_{ij}} = 0$ сводится к решению m несвязанных между собой систем из n линейных уравнений вида

$$\sum_{r=1}^n \frac{\Theta_r \Theta_p}{\sigma^2(y_j)} \xi_{rj} + \frac{\xi_{pj}}{\sigma^2(x_{pj})} = \frac{x_{pj}}{\sigma^2(x_{pj})} + \frac{\Theta_p \cdot y_j}{\sigma^2(y_j)};$$

$$p = 1, 2, \dots, n; \quad j = 1, 2, \dots, m.$$

Полученные новые значения $\hat{\xi}$ должны удовлетворять условию (31). В противном случае те ξ_{ij} , которые выходят за указанные границы, заменяются значениями ближайшей граничной точки. В связи с этим иногда можно ожидать увеличения функционала (30) при новых точных значениях переменных ξ по сравнению с предыдущим шагом итерационного процесса. Это приводит к снижению скорости сходимости процесса и даже к возникновению колебаний. Для устранения таких нежелательных последствий после пересчета ξ те наборы ξ_{ij} , с использованием которых произошло увеличение соответствующих

слагаемых функционала

$$F_j = \sum_{i=1}^n \frac{(x_{ij} - \xi_{ij})^2}{\sigma^2(x_{ij})} + \frac{(y_j - \sum_i \Theta_i \xi_{ij})^2}{\sigma^2(y_j)}$$

по сравнению с предыдущей итерацией, заменяются значениями ξ_{ij} с предыдущего шага.

Применение описанного приема позволило в большинстве случаев увеличить скорость сходимости процесса более чем в 3 раза. Когда же первоначальный процесс не сходился (возникали колебания), решение достигалось за 12–17 итераций.

В конкретных приложениях выбор алгоритма регуляризации зависит от условий задачи. В последующих примерах мы рассмотрим возможности алгоритмов для определения пеленгов двух сигналов в условиях наличия аддитивной помехи. Нас будет интересовать, насколько эффективны алгоритмы относительно разрешающей способности (допустимая разность пеленгов сигналов) и можно ли регистрировать сигналы, мощности которых различаются в 10 раз.

Продемонстрируем результаты модельных расчетов оценки пеленгов для двух излучателей гармонических сигналов. Полагаем, что на детекторы падают плоские волны с известной одинаковой несущей частотой ω_0 . Расстояние между вибраторами равно половине длины падающей волны, помеха подчиняется нормальному закону распределения с нулевым математическим ожиданием и ковариационной матрицей $\mathbf{E} = \sigma^2 \mathbf{I}$. Элементы приемной антенной решетки имеют широкую диаграмму направленности.

Следует отметить, что модель (1) является нелинейной относительно пеленгов θ , что вносит определенные трудности в процедуру пеленгации. Для линеаризации (1) применен метод расширенного базиса. Пусть $\{\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_{N_\theta}\}$ — дискретная сетка пеленгов, соответствующая необходимой предельной точности пеленгации. Представим матрицу \mathbf{A} следующим образом:

$$\mathbf{A}(\theta) = \begin{bmatrix} a(\tilde{\theta}_1) & a(\tilde{\theta}_2) & \dots & a(\tilde{\theta}_{N_\theta}) \end{bmatrix}, \quad \dim[\mathbf{A}(\theta)] = M \times N_\theta. \quad (32)$$

Введем вектор $\mathbf{s}(t)$, $\dim[\mathbf{s}(t)] = N_\theta \times 1$, причем n -й его элемент $s_n(t) = u_k(t)$, если пеленг k -го излучателя равен $\tilde{\theta}_n$; в противном случае $s_n(t) = 0$. Можно записать следующую модель:

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \quad t \in \{1, \dots, L\}. \quad (33)$$

Математическая модель (33) — линейная.

В примерах выбран диапазон пеленгов от 0 до π , введены координаты расширенного базиса через 0,1 рад, т.е. 32 координаты. Другими

словами, определяется интенсивность сигналов, поступивших на антенну под углами от 0 до π рад с шагом 0,1 рад. Таким образом, будет зарегистрировано столько сигналов, сколько их поступит на антенну (в реальных условиях шаг в выборе координат должен определяться релейским разрешением). Соотношение числа пеленгуемых сигналов и числа элементов антенной решетки, зависимость разрешения пеленгов и отношения амплитуд регистрируемых сигналов от числа элементов антенной решетки определяются только уровнем помехи. При гауссовских помехах со средним квадратическим отклонением до 5...10% от амплитуды сигнала достаточно 3-5 элементов для надежной пеленгации сигналов.

Исследованы различные методы регуляризации. Они отличаются разрешающей способностью сигналов, видом решения и его дисперсией. В соответствии со своими свойствами каждый метод имеет свою область применения. Для решения задачи (33) наиболее подходящим с точки зрения разрешения сигналов и минимальных погрешностей определения амплитуд сигналов является метод ℓ_1 -регуляризации, с использованием которого получены приведенные ниже результаты. Они сравнивались с возможностями широко применяемого метода пеленгации MUSIC [4].

Пример 1. Пеленгация двух излучателей с близкими пеленгами (рис. 3, 4). Пеленгация производится в момент времени $t = 0$. Параметры набора детекторов: уравнение выходного сигнала для m -го детектора

$$y_m = \sum_{i=1}^K u_i(t) \exp \{j(\omega_0/c)(m - (M + 1)/2)d \cos \theta_k\},$$

$$m = 1, 2, \dots, M; \quad M = 50; \quad d = \lambda/2.$$

Параметры излучателей: $\omega_0 = 2\pi f_0$, $f_0 = 400$ МГц, $u_1(0) = 10$ мВ, $u_2(0) = 7$ мВ, $\theta_2 - \theta_1 \approx 6^\circ$. Параметры помехи: аддитивная гауссовская помеха $\mathbf{n}(t)$ с математическим ожиданием $E[\mathbf{n}(t)] = 0$ и средним квадратическим отклонением $\sigma = 3$ мВ.

Доверительная вероятность приведенного на рис. 3 доверительного интервала равна 0,95.

На рис. 4 круглые маркеры — точные значения сигналов (согласно модели), квадратные маркеры — значения сигналов с помехой.

При соотношении амплитуд 10:7 и СКО шума 3 мВ MUSIC устойчиво различает пики лишь при разнице пеленгов не менее 0,5 рад.

Пример 2. Пеленгация двух излучателей, сигналы которых сильно отличаются по амплитуде (рис. 5, 6). Пеленгация производится в

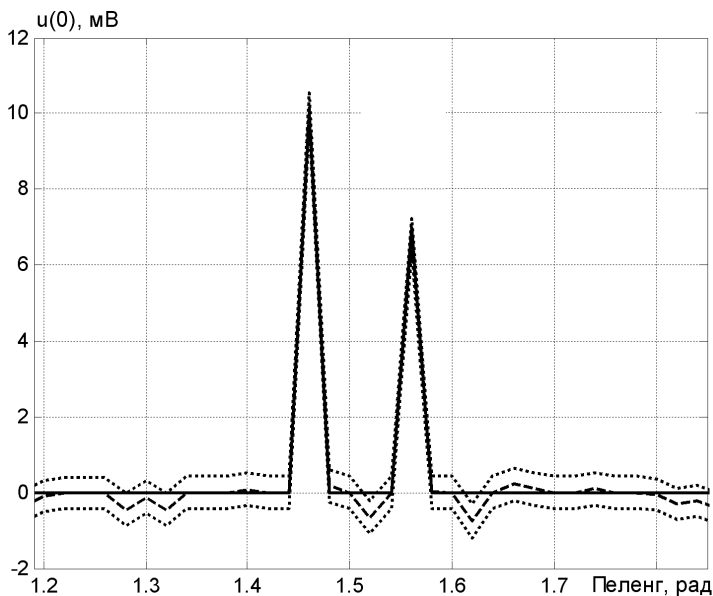


Рис. 3. Результат пеленгации

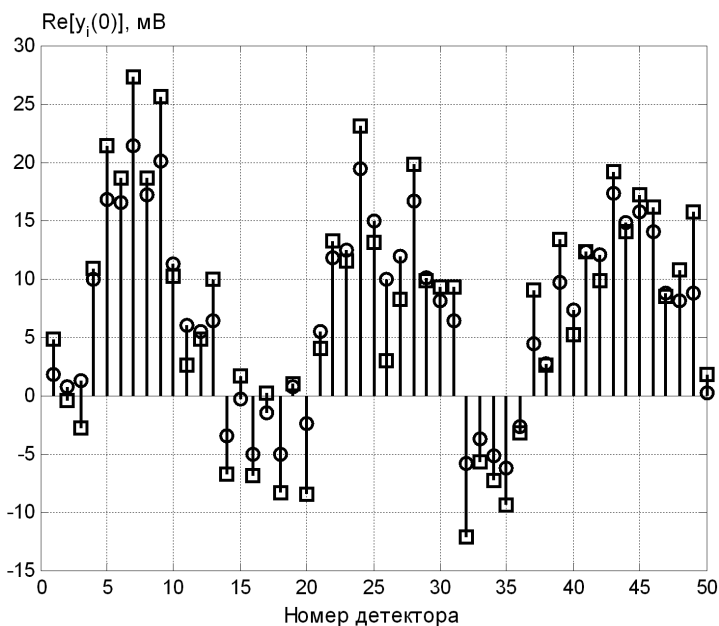


Рис. 4. Сигналы, поступающие с выходов детекторов антенной системы

момент времени $t = 0$. Параметры набора детекторов: уравнение выходного сигнала для m -го детектора

$$y_m = \sum_{i=1}^K u_i(t) \exp \{j(\omega_0/c)(m - (M+1)/2)d \cos \theta_k\}, m = 1, 2, \dots, M;$$

$$M = 12; \quad d = \lambda/2.$$

Параметры излучателей: $\omega_0 = 2\pi f_0$, $f_0 = 400$ МГц, $u_1(0) = 10$ мВ, $u_2(0) = 1$ мВ, $\theta_2 - \theta_1 \approx 11^\circ$. Параметры помехи: аддитивная гауссовская помеха $\mathbf{n}(t)$ с математическим ожиданием $E[\mathbf{n}(t)] = 0$ и средним квадратическим отклонением $\sigma = 2$ мВ.

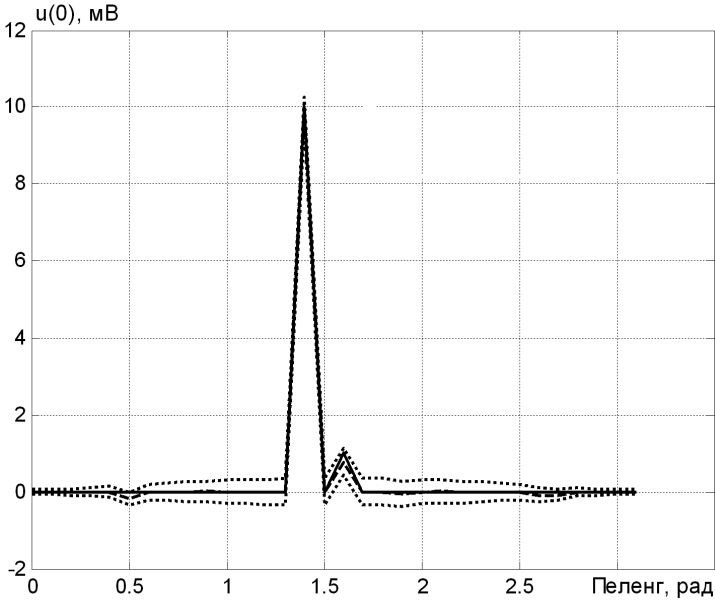


Рис. 5. Результат пеленгации

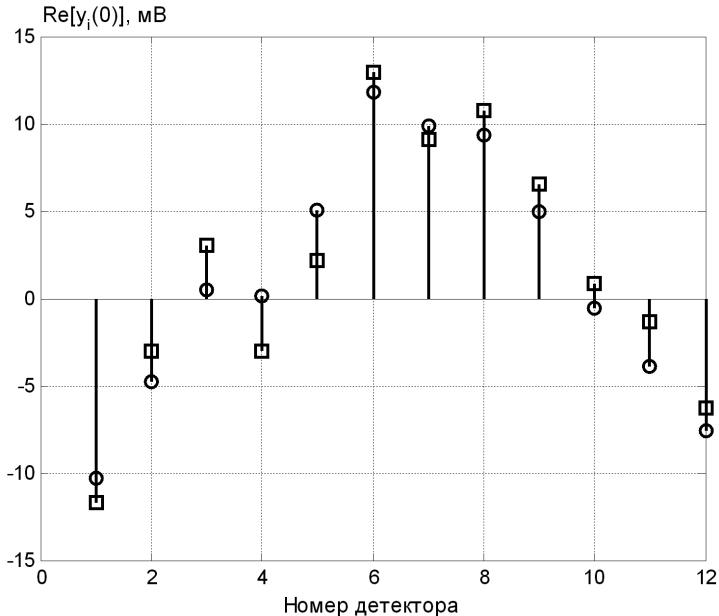


Рис. 6. Сигналы, поступающие с выходов детекторов антенной системы

Доверительная вероятность приведенного на рис. 5 доверительного интервала равна 0,95.

На рис. 6 круглые маркеры — точные значения сигналов (согласно модели), квадратные маркеры — значения сигналов с помехой.

При соотношении амплитуд 10:1 и СКО шума 2 мВ MUSIC устойчиво различает пики лишь при разнице пеленгов не менее 0,8 рад.

Выводы. Исследованы различные алгоритмы методов регуляризации, применяемые в задачах пеленгации источников излучений. Особое внимание уделено методам ℓ_1 - и ℓ_p -регуляризации, которые дают возможность надежно оценивать параметры сигналов, содержащих редкие, резкие импульсы. Предлагаемые алгоритмы позволяют найти не только точечные оценки параметров сигналов, но и ковариационную матрицу оценок (матрицу рассеяния). В примерах демонстрируются пеленги двух гармонических сигналов с постоянной амплитудой. В первом случае обеспечивается разрешение соизмеримых сигналов лучше 6° , во втором сигналы различаются в 10 раз по амплитуде. В обоих случаях пеленги получены методом ℓ_1 -регуляризации с применением переопределенного базиса при высоких уровнях помехи.

СПИСОК ЛИТЕРАТУРЫ

1. Johnson D. H. and Dudgeon D. E. Array signal processing – Concepts and Techniques. Prentice Hall, 1993.
2. Krim H. and Viberg M. Two decades of array signal processing research. The parametric approach // IEEE Signal Proc. Mag., July 1996. – Vol. 13, №. 4. – P. 67–94.
3. Capon J. High resolution frequency-wavenumber spectrum analysis // Proc. IEEE. – 1969. – Vol. 57, №. 8. – P. 1408–1418.
4. Schmidt R. O. A signal subspace approach to multiple emitter location and spectral estimation // Ph.D. thesis, Stanford Univ., 1981.
5. Stoica P. and Sharman K. C. Maximum likelihood methods for direction-of-arrival estimation // IEEE Trans. Signal Processing. July, 1990. – Vol. 38, – №. 7. – P. 1132–1143.
6. Stoica P. and Nehorai A. MUSIC, maximum likelihood, and Cramer-Rao bound // IEEE Trans. May, 1989. ASSP. – Vol. 37. – №. 5. – P. 720–741
7. Engl H. W., Hanneke M. and Neubauber A. Regularization of Inverse Problems. Kluwer, 1996.
8. Neumaier A. Solving ill-conditioned and singular linear systems: a tutorial on regularization. // SIAM Review. – 1998. – Vol. 40, – №. 3. – P. 636–666.
9. Hansen P. C. Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems // Numer. Algorithms. – 1994. – Vol. 6. – P. 1–35.
10. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. – М.: Наука, 1979.
11. Karl W. C. Regularization in image restoration and reconstruction // Handbook of Image and Video Processing. A. Bovik. Ed. Academic Press, 2000.
12. Грешилов А. А. Некорректные задачи цифровой обработки информации и сигналов. – М.: Радио и связь, 1984.

13. Stoica P. and Nehorai A. MUSIC, maximum likelihood, and Cramer-Rao bound // IEEE Trans. ASSP. May, 1989. – Vol. 37. – №. 5. – P. 720–741.
14. Грешилов А. А. Математические методы принятия решений: Учеб. пособие для вузов. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2006.
15. Mallat S. A wavelet tour of signal processing. – Academic Press, 1998.
16. Donoho D. L., Johnstone I. M., Koch J. C., Stern A. S. Maximum entropy and the nearly black object // J. R. Statist. Soc. B. – 1992. – Vol. 54, №. 1. – P. 41–81.
17. Gorodnitsky I. F. and Rao B. D. Sparse signal reconstruction from limited data using focus: a re-weighted minimum norm algorithm // IEEE Trans. Signal Processing. Mar. 1997. – Vol. 45, №. 3. – P. 600–616.
18. Chen S. S., Donoho D. L. and Saunders M. A. Atomic decomposition by basis pursuit // SIAM J. Scientific Computing. – 1998. – Vol. 20, №. 1. – P. 33–61.
19. Fu W. J. Penalized regressions: the bridge versus the LASSO // Journal of Computational and Graphical Statistics. – Sept. 1998. – Vol. 7, №. 3.
20. Geman D. and Yang C. Nonlinear image recovery with half-quadratic regularization // IEEE Trans. Image Processing. – July, 1995. – Vol. 4, №. 7. P. 932–946.
21. Miller A. J. Subset Selection in Regression. – Chapman and Hall, 2002.
22. Tibshirani R. Regression shrinkage and selection via the LASSO // Journal of Royal Statistical Society, Series B. – Nov., 1996. – Vol. 58. – P. 267–288.
23. Cetin M. and Karl W. C. Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization // IEEE Trans. Image Processing. – Apr. 2001. – Vol. 10. – №. 4. – P. 623–631.
24. Mangasarian O. L. and Musicant D. R. Robust linear and support vector regression // IEEE Trans. Pattern Anal. Machine Intell. – 2000. – Vol. 22, №. 9. – P. 950–955.
25. Vogel C. R. and Oman M. E. Fast, robust total variation-based reconstruction of noisy, blurred images // IEEE Trans. Image Processing. Jun, 1998. – Vol. 7, №. 6. – P. 813–824.
26. Geman D. and Reynolds G. Constrained restoration and the recovery of discontinuities // IEEE Trans. Pattern Anal. Machine Intell. Mar., 1992. – Vol. 14, №. 3. P. 367–383.
27. Hoerl A. E. and Kennard R. W. Ridge regression: Biased estimation for nonorthogonal problems // Technometrics. – 1970. – Vol. 12, №. 3. – P. 55–67.
28. Charbonnier P., Blanc-Feraud L., Aubert G. and Barlaud M. Deterministic edge-preserving regularization in computed imaging // IEEE Trans. Image Processing. – Feb. 1997. – Vol. 6, №. 2. – P. 298–310.
29. Delaney A. H. and Bresler Y. Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography // IEEE Trans. Image Processing. – Feb., 1998. – Vol. 7, №. 2. – P. 204–221.
30. Malioutov D. M. A sparse signal reconstruction perspective for source localization with sensor arrays // Master of Science thesis, Massachusetts Institute of Technology, 2003.

Статья поступила в редакцию 19.01.2007

Анатолий Антонович Грешилов родился в 1939 г., окончил в 1964 г. Московский инженерно-физический институт. Д-р техн. наук, профессор кафедры “Вычислительная математика и математическая физика” МГТУ им. Н.Э. Баумана. Автор более 150 научных работ, в том числе 19 монографий, 20 авторских свидетельств и патентов в области разработки математических методов строгого учета неопределенности исходной информации в задачах математической физики, распознавания образов, прогнозирования и других технических приложений.



A.A. Greshilov (b. 1939) graduated from the Moscow Institute for Engineering and Physics in 1964. D. Sc. (Eng.) professor of “Computational Mathematics and Mathematical Physics” department of the Bauman Moscow State Technical University. Author of more than 150 publications including 19 monographs, 20 author’s certificates and patents in the field of development of mathematical methods for strict account of source data uncertainty in problems of mathematical physics, image identification, forecast and other technical applications.

Борис Петрович Назаренко родился в 1943 г., окончил в 1966 г. МВТУ им. Н.Э. Баумана. Канд. техн. наук, руководитель научно-учебного комплекса, декан факультета “Фундаментальные науки” МГТУ им. Н.Э. Баумана. Автор более 70 научных работ, в том числе трех монографий и трех авторских свидетельств в области математического моделирования транспортных систем, теоретической механики и технических приложений информатики.



B.P. Nazarenko (b. 1943) graduated from the Bauman Moscow Higher Technical School in 1966. Ph. D. (Eng.), chief of scientific and educational complex, dean of “Fundamental Sciences” faculty of the Bauman Moscow State Technical University. Author of more than 70 publications including 3 monographs and 3 author’s certificates in the field of mathematical simulation of transport systems, theoretical mechanics and technical applications of information technology.

Павел Анатольевич Плохута родился в 1983 г., окончил МГТУ им. Н.Э. Баумана в 2006 г. Аспирант кафедры “Вычислительная математика и математическая физика” МГТУ им. Н.Э. Баумана.



P.A. Plokhuta (b. 1983) graduated from the Bauman Moscow State Technical University in 2006. Post-graduate of “Computational Mathematics and Mathematical Physics” department of the Bauman Moscow State Technical University.