УДК 519.8
П.И.СТЕЦЮК, А.В.МИЦА

О ЗАДАЧАХ ОПТИМИЗАЦИИ ПАРАМЕТРОВ ДЛЯ МНОГОСЛОЙНЫХ ОПТИЧЕСКИХ ПОКРЫТИЙ

Ключевые слова: многослойное оптическое покрытие, характеристическая матрица, коэффициент пропускания излучения, задача нелинейного программирования, субградиентный метод.

В работе рассмотрены две задачи для нахождения оптимальных параметров плоских многослойных оптических покрытий, которые сформулированы как многоэкстремальные задачи нелинейного программирования со сложным видом целевой функции. Обсуждаются вопросы нахождения локальных экстремумов этих задач с помощью методов первого порядка и анализируются способы вычисления градиента целевой функции в зависимости от количества слоев в оптическом покрытии.

1 Основные обозначения

Пусть некоторое плоское многослойное оптическое покрытие (МОП) состоит из N плоских однородных слоев. Каждый k-й слой ($k = 1, 2, \ldots, N$) характеризуется конкретными значениями коэффициента преломления n_k и геометрической толщины d_k. Параметры МОП обозначим:

$$\vec{n} = \begin{pmatrix} n_1 \\ \vdots \\ n_N \end{pmatrix}, \quad \vec{d} = \begin{pmatrix} d_1 \\ \vdots \\ d_N \end{pmatrix},$$

где $\vec{n}$ – вектор значений коэффициентов преломления плоских слоев (n_k, $k = 1, \ldots, N$), а $\vec{d}$ – вектор значений геометрических толщин этих же слоев (d_k, $k = 1, \ldots, N$).

Пусть электромагнитная волна длины λ падает перпендикулярно поверхности МОП и ее энергия не теряется при прохождении через МОП. Характеристическая матрица для МОП равна произведению характеристических матриц всех N плоских слоев, т.е.

$$M(\vec{n}, \vec{d}, \lambda) = \prod_{k=1}^{N} M(n_k, d_k, \lambda), \tag{1}$$

где $M(n_k, d_k, \lambda)$ – характеристическая (2×2)-матрица k-го слоя ($k = 1, \ldots, N$) и имеет следующий вид:

$$M(n_k, d_k, \lambda) = \left\| \begin{array}{cc} \cos\left(\frac{2\pi n_k d_k}{\lambda}\right) & -\frac{i}{n_k}\sin\left(\frac{2\pi n_k d_k}{\lambda}\right) \\ -in_k \sin\left(\frac{2\pi n_k d_k}{\lambda}\right) & \cos\left(\frac{2\pi n_k d_k}{\lambda}\right) \end{array} \right\|. \tag{2}$$

Здесь i – "мнимая"единица.

Формула (1) лежит в основе метода Абелле для расчета параметров оптической многослойной структуры (прямая задача) [1]. Она дает возможность рассчитать характеристическую матрицу для МОП, зная характеристические матрицы для всех плоских однородных слоев, из которых состоит эта МОП. В результате расчета по формуле (1) характеристическая матрица $M(\vec{n}, \vec{d}, \lambda)$ будет иметь следующий вид:

$$M(\vec{n}, \vec{d}, \lambda) = \left\| \begin{array}{cc} m_{11}(\vec{n}, \vec{d}, \lambda) & im_{12}(\vec{n}, \vec{d}, \lambda) \\ im_{21}(\vec{n}, \vec{d}, \lambda) & m_{22}(\vec{n}, \vec{d}, \lambda) \end{array} \right\| \equiv \left\| \begin{array}{cc} m_{11} & im_{12} \\ im_{21} & m_{22} \end{array} \right\|.$$

Отметим, что характеристические матрицы вида (2) удовлетворяют следующему условию:

$$\det(M(n_k, d_k, \lambda)) = 1, \quad k = 1, 2, \ldots, N,$$

что гарантирует выполнение такого же условия для характеристической матрицы МОП, т.е.

$$\det(M(\vec{n}, \vec{d}, \lambda)) = 1.$$

Данное свойство имеет простой физический смысл. Оно означает, что если электромагнитная волна распространяется в N непоглощающих ее энергию средах, то любая комбинированная (из этих N сред) среда также не будет поглощать энергию электромагнитной волны.

Коэффициент пропускания энергии электромагнитной волны длины λ через МОП зависит от коэффициентов характеристической матрицы $M(\vec{n}, \vec{d}, \lambda)$ и имеет следующий вид:

$$T(\vec{n}, \vec{d}, \lambda) = \frac{4}{2 + \frac{n_0}{n_s} m_{11}^2 + \frac{n_s}{n_0} m_{22}^2 + n_0 n_s m_{12}^2 + \frac{1}{n_0 n_s} m_{21}^2} \qquad (3)$$

где n_0 и n_s – показатели преломления внешней среды и подкладки (соответственно). Коэффициент $T(\vec{n}, \vec{d}, \lambda)$, вычисленный согласно формулы (3), означает, что проходящая через МОП, которая ограничена снизу подкладкой с показателем преломления n_s, енергия электромагнитной волны длины λ будет равна

$$E(\lambda) = T(\vec{n}, \vec{d}, \lambda) * E_0(\lambda).$$

где $E_0(\lambda)$ – энергия падающей на МОП (перпендикулярно к ее поверхности) электромагнитной волны из внешней среды с показателем преломления n_0.

Следовательно, зная для МОП параметры $\vec{n}$ и $\vec{d}$ мы можем по формуле (3) рассчитать коэффициент пропускания энергии электромагнитной волны $T(\vec{n}, \vec{d}, \lambda)$ при заданной длине волны λ, который гарантируется на выходе из МОП. Оптимизационные задачи, которые рассмотрим в следующем разделе, будут предназначены для решения "обратных"задач, т.е. для заданной структуры МОП будем подбирать такие ее параметры $\vec{n}$ и $\vec{d}$, чтобы на выходе из МОП наилучшим образом обеспечивалось функционирование заданных коэффициентов пропускания излучения для некоторого дискретного спектра длин волн.

2 Две задачи оптимизации параметров МОП

Пусть задан дискретный спектр длин волн λ_j, $j = 1, \ldots, L$, и сответствующие этим длинам волн коэффициенты пропускания излучения: T_j, $j = 1, \ldots, L$, которые требуется обеспечить на выходе из МОП. Параметры внешней оболочки для МОП, т.е. n_0 – показатель преломления внешней среды и n_s – показатель преломления подкладки считаются фиксированными.

Пусть $\vec{n}^*$ – вектор оптимальных значений коэффициентов преломления для МОП, а $\vec{d}^*$ – вектор оптимальных значений геометрических толщин.

Первая оптимизационная задача (задача A) связана с нахождением таких параметров $\vec{n}^*$ и $\vec{d}^*$, которые для заданного спектра длин волн позволяют обеспечить оптимальные (по критерию наилучшего среднеквадратичного отклонения) коэффициенты прохождения излучения через МОП при двухсторонних ограничениях на значения параметров $\vec{n}^*$ и $\vec{d}^*$. Оптимальные параметры $\vec{n}^*$ и $\vec{d}^*$ находятся посредством решения следующей задачи нелинейного программирования:

$$\min_{\vec{n},\vec{d}\in R^N} \left\{ F(\vec{n},\vec{d}) = \sum_{j=1}^{L} \Big(T_j - T(\vec{n},\vec{d},\lambda_j) \Big)^2 \right\} \tag{1A}$$

при ограничениях:

$$\vec{n}_{low} \le \vec{n} \le \vec{n}_{up}, \tag{2A}$$

$$\vec{d}_{low} \le \vec{d} \le \vec{d}_{up}. \tag{3A}$$

Здесь $\vec{n}_{low}$ и $\vec{n}_{up}$ – векторы нижних и верхних границ (соответственно) на коэффициенты преломления слоев; $\vec{d}_{low}$ и $\vec{d}_{up}$ – векторы нижних и верхних границ (соответственно) на геометрические толщины плоских слоев.

Вторая оптимизационная задача (задача B) связана с нахождением оптимальных параметров $\vec{n}^*$ и $\vec{d}^*$ при заданной геометрической толщине МОП и двухсторонних ограничениях на неизвестные параметры $\vec{n}^*$. Задача B формулируется следующей задачей нелинейного программирования:

$$\min_{\vec{n},\vec{d}\in R^N} \left\{ F(\vec{n},\vec{d}) = \sum_{j=1}^{L} \Big(T_j - T(\vec{n},\vec{d},\lambda_j) \Big)^2 \right\} \tag{1B}$$

при ограничениях:

$$\vec{n}_{low} \le \vec{n} \le \vec{n}_{up} \tag{2B}$$

$$\sum_{k=1}^{N} d_k = d_{total} \tag{3B}$$

$$d_k \ge 0, \quad k = 1, \ldots, N, \tag{4B}$$

где d_{total} – суммарная геометрическая толщина МОП. Целевая функция для задачи B такая же как и для задачи A, но другая система ограничений, которая требует выдержать заданную геометрическую толщину МОП при двухсторонних ограничениях на неизвестные коэффициенты преломления $\vec{n}^*$.

С помощью задачи B можно изучать влияние неоднородности показателя преломления посредством некоторого процесса линеаризации неоднородных оптических структур (т.е. рассмотрения неоднородной структуры как некоторого набора однородных слоев). Так, например, рассматривая неоднородную оптическую структуру как многослойную структуру с заданными значениями коэффициентов преломления для каждого однородного слоя, можно найти "оптимальные"значения толщин этих слоев при ограничении на их суммарную геометрическую толщину. Найденное решение может помочь при исследовании тех неизвестных законов (например, ступенчатый, линейный или экспоненциальный) [1], по которым изменяются коэффициенты преломления неоднородных оптических покрытий в направлении, которое перпендикулярно границам слоев.

Задачи A и B являются многоэкстремальными задачами нелинейного программирования с количеством переменных $2N$. Их характеризуют сложная функция цели и достаточно простые системы ограничений. Для нахождения локальных экстремумов оптимизационных задач A и B можно использовать методы первого порядка, т.е. методы, которые используют вычисленные в точках итерационного процесса значения функции $F(\vec{n}, \vec{d})$ и ее градиента. В качестве таковых можно использовать различные градиентные и субградиентные методы [2]-[9]. Так, например, в [1, 10] эффективными методами первого порядка при решении подобных задач указываются методы из семейства квазиньютоновских методов и сопряженных направлений [4,9].

Однако, для нахождения локальных экстремумов задач A и B целесообразнее использовать эффективные варианты субградиентных методов с растяжением пространства (например, r-алгоритмы) [2]-[3]. В пользу этого можно привести два довода. Во-первых, применение r-алгоритмов позволяет избежать использования негладких штрафных функций для учета ограничений (2A)-(3A) и (2B)-(4B). Во-вторых, при использовании r-алгоритмов легко расширить класс критериев оптимальности для параметров $\vec{n}^*$ и $\vec{d}^*$ в задачах A и B. Так, например, заменив функцию $F(\vec{n}, \vec{d})$ на

$$F_1(\vec{n}, \vec{d}) = \sum_{j=1}^{L} \left| T_j - T(\vec{n}, \vec{d}, \lambda_j) \right|,$$

где $|\cdot|$ – модуль (абсолютная величина) числа, получим наилучшие значения $\vec{n}^*$ и $\vec{d}^*$ согласно критерию наименьших модулей. Чтобы выбрать наилучшие параметры согласно минимаксному чебышевскому критерию достаточно заменить $F(\vec{n}, \vec{d})$ на

$$F_2(\vec{n}, \vec{d}) = \max_{j=1,\dots,L} \left| T_j - T(\vec{n}, \vec{d}, \lambda_j) \right|.$$

Две последних – негладкие функции, и для работы с ними требуются именно субградиентные методы.

Использование градиентного (субградиентного) метода подразумевает наличие некоторой процедуры (оракула), которая вычисляет значение функции $F(\vec{n},\vec{d})$ и ее градиента (субградиентов функций $F_1(\vec{n},\vec{d})$ и $F_2(\vec{n},\vec{d})$) в заданной точке. Основные вычислительные затраты при подготовке оракула для задач A и B связаны с вычислением градиентов функций $T(\vec{n},\vec{d},\lambda_j)$ для всех $j=1,2,\ldots,L$.

Эффективное построение оракула (по количеству вычислений) будет определять скорость работы того или иного выбранного метода первого порядка. Поэтому в следующих разделах рассмотрим и проанализируем два способа (конечно-разностный и аналитический) для вычисления градиента функции $F(\vec{n},\vec{d})$ и проведем анализ компьтерных реализаций обоих способов в зависимости от количества слоев N. Оба обсуждаемых способа можно также использовать и для расчета субградиентов негладких функций $F_1(\vec{n},\vec{d})$ и $F_2(\vec{n},\vec{d})$.

3 Способы вычисления градиента $F(\vec{n},\vec{d})$

В разделе проанализируем два способа вычисления градиента функции $F(\vec{n},\vec{d})$: (i) конечно-разностный и (ii) аналитический способы. При этом отметим как недостатки, так и преимущества обоих способов. Компоненты вектор-градиента $\nabla F(\vec{n}_k,\vec{d}_k)$, вычисленного в точке $(\vec{n}_k,\vec{d}_k)$, условимся обозначать $\nabla_i F(\vec{n}_k,\vec{d}_k)$, $i=1,\ldots,2N$. Для упрощения обозначений индекс k будем опускать там, где это не приводит к недоразумению.

3.1 Конечно-разностный способ.

Пусть $\vec{e}_i$ – N-мерный вектор, i-ая компонента которого равна единице, а все остальные компоненты равны нулю.

Задаемся достаточно малыми величинами Δn_i и Δd_i, $i=1,\ldots,N$. Тогда компоненты градиента функции $F(\vec{n},\vec{d})$ в точке $(\vec{n},\vec{d})$ рассчитываются по следующим формулам:

$$\begin{aligned}
\nabla_i F(\vec{n}_k,\vec{d}_k) &= \tfrac{F(\vec{n}_k+\Delta n_i\vec{e}_i,\vec{d}_k)-F(\vec{n}_k,\vec{d}_k)}{\Delta n_i}, i=1,\ldots,N,\\
\nabla_{N+i} F(\vec{n}_k,\vec{d}_k) &= \tfrac{F(\vec{n}_k,\vec{d}_k+\Delta d_i\vec{e}_i)-F(\vec{n}_k,\vec{d}_k)}{\Delta d_i}, i=1,\ldots,N.
\end{aligned}$$

Хотя это и самый простой для реализации способ вычисления $\nabla F(\vec{n},\vec{d})$, он обладает двумя существенными недостатками:

I) разномасштабность параметров $\vec{n}$ и $\vec{d}$ (в реальных задачах коэффициенты преломления составляют несколько единиц, а геометрические толщины порядка нескольких сотен нанометров (10^{-9}м.)) требует достаточно аккуратного обращения с величинами Δn_i и Δd_i, $i=1,\ldots,N$, при построении правил остановки итерационного процесса.

II) точность нахождения локального экстремума существенно зависит от используемых значений Δn_i и Δd_i, $i=1,\ldots,N$. Это означает, что попытки найти оптимальные параметры $\vec{n}^*$ и $\vec{d}^*$ с точностью близкой к той точности, с которой вычисляются градиенты в точках, заранее обречены на неудачу. Более того, попытка осуществить такой расчет будет характеризоваться расходимостью субградиентного процесса в окрестности локального экстремума, и связано это будет с тем, что из-за достаточно "грубого"вычисления градиента выбираются неправильные направления движения для субградиентного процесса.

3.2 Аналитический способ

Аналитический способ вычисления градиента функции $F(\vec{n}, \vec{d})$ свободен от недостатков, свойственных конечно-разностному способу. Он состоит в следующем. Непосредственное дифференцирование функции $F(\vec{n}, \vec{d})$ дает следующие формулы для вычисления компонент ее градиента в точке $(\vec{n}_k, \vec{d}_k)$:

$$\begin{aligned}
\nabla_i F(\vec{n}_k, \vec{d}_k) &= 2\sum_{j=1}^{L}\Big(T_j - T(\vec{n}_k, \vec{d}_k, \lambda_j)\Big)\tfrac{\partial T(\vec{n}_k, \vec{d}_k, \lambda_j)}{\partial n_i}, i = 1, \dots, N, \\
\nabla_{N+i} F(\vec{n}_k, \vec{d}_k) &= 2\sum_{j=1}^{L}\Big(T_j - T(\vec{n}_k, \vec{d}_k, \lambda_j)\Big)\tfrac{\partial T(\vec{n}_k, \vec{d}_k, \lambda_j)}{\partial d_i}, i = 1, \dots, N.
\end{aligned}$$

Следовательно, для аналитического вычисления $\nabla F(\vec{n}_k, \vec{d}_k)$ достаточно для всех $i = 1, \dots, N$ уметь находить аналитические выражения для производных $\frac{\partial T(\vec{n}, \vec{d}, \lambda)}{\partial n_i}$ и $\frac{\partial T(\vec{n}, \vec{d}, \lambda)}{\partial d_i}$ при фиксированном значении λ. Используя (3) эти производные легко найти для каждого i $(1 \le i \le N)$ по следующему правилу:

$$\frac{\partial T(\vec{n}, \vec{d}, \lambda)}{\partial n_i} = \frac{8\left(\frac{n_0}{n_s} m_{11}\frac{\partial m_{11}}{\partial n_i} + \frac{n_s}{n_0} m_{22}\frac{\partial m_{22}}{\partial n_i} + n_0 n_s m_{12}\frac{\partial m_{12}}{\partial n_i} + \frac{1}{n_0 n_s} m_{21}\frac{\partial m_{21}}{\partial n_i}\right)}{\left(2 + \frac{n_0}{n_s} m_{11}^2 + \frac{n_s}{n_0} m_{22}^2 + n_0 n_s m_{12}^2 + \frac{1}{n_0 n_s} m_{21}^2\right)^2},$$

$$\frac{\partial T(\vec{n}, \vec{d}, \lambda)}{\partial d_i} = \frac{8\left(\frac{n_0}{n_s} m_{11}\frac{\partial m_{11}}{\partial d_i} + \frac{n_s}{n_0} m_{22}\frac{\partial m_{22}}{\partial d_i} + n_0 n_s m_{12}\frac{\partial m_{12}}{\partial d_i} + \frac{1}{n_0 n_s} m_{21}\frac{\partial m_{21}}{\partial d_i}\right)}{\left(2 + \frac{n_0}{n_s} m_{11}^2 + \frac{n_s}{n_0} m_{22}^2 + n_0 n_s m_{12}^2 + \frac{1}{n_0 n_s} m_{21}^2\right)^2},$$

где $\frac{\partial m_{11}}{\partial n_i}$, $\frac{\partial m_{12}}{\partial n_i}$, $\frac{\partial m_{21}}{\partial n_i}$, $\frac{\partial m_{22}}{\partial n_i}$, $\frac{\partial m_{11}}{\partial d_i}$, $\frac{\partial m_{12}}{\partial d_i}$, $\frac{\partial m_{21}}{\partial d_i}$ и $\frac{\partial m_{22}}{\partial d_i}$ – коэффициенты следующих (2×2)-матриц:

$$\frac{\partial M(\vec{n}, \vec{d}, \lambda)}{\partial n_i} = \begin{Vmatrix} \frac{\partial m_{11}}{\partial n_i} & i\frac{\partial m_{12}}{\partial n_i} \\ i\frac{\partial m_{21}}{\partial n_i} & \frac{\partial m_{22}}{\partial n_i} \end{Vmatrix}; \quad \frac{\partial M(\vec{n}, \vec{d}, \lambda)}{\partial d_i} = \begin{Vmatrix} \frac{\partial m_{11}}{\partial d_i} & i\frac{\partial m_{12}}{\partial d_i} \\ i\frac{\partial m_{21}}{\partial d_i} & \frac{\partial m_{22}}{\partial d_i} \end{Vmatrix}.$$

Матрицы $\frac{\partial M(\vec{n}, \vec{d}, \lambda)}{\partial n_i}$ и $\frac{\partial M(\vec{n}, \vec{d}, \lambda)}{\partial d_i}$ легко рассчитать используя соотношения (1) и (2). Так, например, матрицы $\frac{\partial M(\vec{n}, \vec{d}, \lambda)}{\partial d_i}$ рассчитываются с помощью следующих соотношений:

$$\begin{array}{l}
\frac{\partial M(\vec{n},\vec{d},\lambda)}{\partial d_1} = \frac{\partial M(n_1,d_1,\lambda)}{\partial d_1}\prod\limits_{k=2}^{N} M(n_k,d_k,\lambda),\\
\frac{\partial M(\vec{n},\vec{d},\lambda)}{\partial d_i} = \prod\limits_{k=1}^{i-1} M(n_k,d_k,\lambda)\frac{\partial M(n_i,d_i,\lambda)}{\partial d_i}\prod\limits_{k=i+1}^{N} M(n_k,d_k,\lambda), \quad i=2,\dots,N-1,\\
\frac{\partial M(\vec{n},\vec{d},\lambda)}{\partial d_N} = \prod\limits_{k=1}^{N-1} M(n_k,d_k,\lambda)\frac{\partial M(n_N,d_N,\lambda)}{\partial d_N},
\end{array} \tag{4}$$

где (2×2)-матрицы $\frac{\partial M(n_i,d_i,\lambda)}{\partial d_i}$ для каждого $i=1,\dots,N$ имеют следующий вид:

$$\frac{\partial M(n_i,d_i,\lambda)}{\partial d_i} = \left\|\begin{array}{cc} \frac{\partial}{\partial d_k}\left(\cos\left(\frac{2\pi n_k d_k}{\lambda_j}\right)\right) & -i\frac{\partial}{\partial d_k}\left(\frac{1}{n_k}\sin\left(\frac{2\pi n_k d_k}{\lambda_j}\right)\right) \\ -i\frac{\partial}{\partial d_k}\left(n_k\sin\left(\frac{2\pi n_k d_k}{\lambda_j}\right)\right) & \frac{\partial}{\partial d_k}\left(\cos\left(\frac{2\pi n_k d_k}{\lambda_j}\right)\right)\end{array}\right\| =$$

$$= \left\|\begin{array}{cc} -\frac{2\pi n_i}{\lambda_j}\sin\left(\frac{2\pi n_i d_i}{\lambda_j}\right) & -\frac{i}{n_i}\frac{2\pi n_i}{\lambda_j}\cos\left(\frac{2\pi n_i d_i}{\lambda_j}\right) \\ -in_i\frac{2\pi n_i}{\lambda_j}\cos\left(\frac{2\pi n_i d_i}{\lambda_j}\right) & \\ & -\frac{2\pi n_i}{\lambda_j}\sin\left(\frac{2\pi n_i d_i}{\lambda_j}\right)\end{array}\right\|.$$

Аналогично рассчитываются и матрицы $\frac{\partial M(\vec{n},\vec{d},\lambda)}{\partial n_i}$. Явный вид для них опустим в силу громоздкости аналитического выражения для матриц

$$\frac{\partial M(n_i,d_i,\lambda)}{\partial n_i} = \left\|\begin{array}{cc} \frac{\partial}{\partial n_k}\left(\cos\left(\frac{2\pi n_k d_k}{\lambda_j}\right)\right) & -i\frac{\partial}{\partial n_k}\left(\frac{1}{n_k}\sin\left(\frac{2\pi n_k d_k}{\lambda_j}\right)\right) \\ -i\frac{\partial}{\partial n_k}\left(n_k\sin\left(\frac{2\pi n_k d_k}{\lambda_j}\right)\right) & \frac{\partial}{\partial n_k}\left(\cos\left(\frac{2\pi n_k d_k}{\lambda_j}\right)\right)\end{array}\right\|.$$

Аналитический способ позволяет вычислять точное (в пределах компьютерной арифметики) значение градиента функции $F(\vec{n},\vec{d})$ в точке $(\vec{n}_k,\vec{d}_k)$ и по затратам арифметических операций он такой же как и конечно-разностный способ. Аналитический способ целесообразней использовать, чем конечно-разностный, когда требуется высокая точность по нахождению оптимальных параметров $(\vec{n}^*,\vec{d}^*)$.

4 О двух реализациях вычисления $\nabla F(\vec{n},\vec{d})$

Основной параметр, от которого зависит скорость сходимости методов первого порядка, есть размер N (количество слоев в МОП). В зависимости от типа задач количество слоев может составлять от нескольких единиц до нескольких десятков. Так, например, в задачах, связанных с "отрезающими"фильтрами, их количество может составлять от 20 до 50. В силу этого смысл приобретают процедуры быстрого вычисления градиента функции $F(\vec{n}, \vec{d})$.

Далее рассмотрим две возможных реализации для вычисления $\nabla F(\vec{n}, \vec{d})$: прямолинейная и ускоренная реализации. Обе реализации могут быть использованы как при конечно-разностном, так и при аналитическом способах вычисления градиента. При ускоренной реализации убираются (за счет использования дополнительной памяти для запоминания промежуточных результатов) повторяющиеся умножения матриц для различных $i = 1, \ldots, N$, которые имеют место при прямолинейной реализации. Анализ ускорения для обоих способов показывает, что по умножению матриц ускоренная реализация выигрывает у прямолинейной реализации и этот выиграш растет с ростом N. Его целесообразно использовать при работе с количеством слоев порядка 20-40.

Содержательный смысл того ускорения, которое можно использовать при вычислении градиентов для обоих описанных способов, объясним на примере вычисления $\frac{\partial M(\vec{n},\vec{d},\lambda)}{\partial d_i}$ по формуле (4). Рассмотрим две следующих реализации вычисления матриц $\frac{\partial M(\vec{n},\vec{d},\lambda)}{\partial d_i}$ для всех $i = 1, \ldots, N$:

(i) Прямолинейная реализация, в основе которой лежит непосредственное применение формулы (4). Для каждого $i = 1, \ldots, N$ она требует $(N-1)$ умножений матриц размером 2×2. Следовательно, прямолинейная реализация потребует

$$N_1^* = N(N-1)$$

умножений матриц размером 2×2 (т.е. N раз требуется использовать $(N-1)$ умножение матриц).

(ii) Ускоренная реализация, при которой убираются повторяющиеся умножения матриц для различных $i = 1, \ldots, N$, свойственные прямолинейной реализации. Суть ускоренной реализации состоит в следующем. Один раз вычислим такие матрицы:

$$\begin{aligned} M_i^+ &= \prod_{k=1}^{i} M(n_k, d_k, \lambda_j), \quad i = 1, \ldots, N-1, \\ M_i^- &= \prod_{k=i}^{N} M(n_k, d_k, \lambda_j), \quad i = 2, \ldots, N, \end{aligned}$$

для чего достаточно $2(N-2)$ операций умножения матриц размером 2×2. Затем для вычисления матриц $\frac{\partial M(\vec{n},\vec{d},\lambda)}{\partial d_i}$ для всех $i = 1, \ldots, N$ взамен формулы (4) будем использовать следующую формулу:

$$\begin{aligned} \frac{\partial M(\vec{n},\vec{d},\lambda_j)}{\partial d_1} &= \frac{\partial M(n_1,d_1,\lambda_j)}{\partial d_1} * M_2^-, \\ \frac{\partial M(\vec{n},\vec{d},\lambda_j)}{\partial d_i} &= M_{i-1}^+ * \frac{\partial M(n_i,d_i,\lambda_j)}{\partial d_i} * M_{i+1}^-, i = 2, \ldots, N-1 \ , \\ \frac{\partial M(\vec{n},\vec{d},\lambda_j)}{\partial d_N} &= M_{N-1}^- * \frac{\partial M(n_N,d_N,\lambda_j)}{\partial d_N}. \end{aligned} \tag{4a}$$

которая есть та же формула (4), но записана в более компактном для практического применения виде. Для реализации формулы (4а) требуется $2+2(N-2)=2(N-1)$ умножений матриц размером 2×2. Следовательно, для вычисления матриц $\frac{\partial M(\vec{n},\vec{d},\lambda)}{\partial d_i}$ для всех $i=1,\ldots,N$ ускоренная реализация потребует

$$N_2^* = 2(N-2)+2(N-1) = 2(2N-3)$$

умножений матриц размером 2×2.

По умножению матриц ускоренная реализация выигрывает у прямолинейной реализации и этот выиграш будет характеризовать следующая величина:

$$q(N) = \frac{N_2^*}{N_1^*} = \frac{2(2N-3)}{N(N-1)} = \frac{2}{N} + \frac{2(N-2)}{N(N-1)}.$$

При $N=2$ и $N=3$ имеем $q(2)=1$ и $q(3)=1$, что означает, что при количестве слоев, не превышающем трех, все равно какую из реализаций использовать для вычисления градиентов. Однако, уже при $N=4$ имеется незначительный выигрыш

$$q(4) = \frac{2}{4} + \frac{2\cdot 2}{4\cdot 3} = \frac{1}{2} + \frac{1}{3} = \frac{5}{6},$$

который возрастает с ростом N. Так, например, при $N=20$ имеем

$$q(20) = \frac{1}{10} + \frac{2\cdot 18}{20\cdot 19} \approx \frac{1}{5},$$

что означает, что при работе с двадцати слойными МОП можно сэкономить в пять раз на вычислениях, связанных с умножением матриц размером 2×2. Еще большая экономия (в 10 раз) получается при работе с 40-слойными МОП. Однако следует отметить, что при решении задач А и В с $N=20\div 40$ мы не получим "чистого"выиграша по времени от пяти до десяти раз. Дело в том, что расчет характеристических матриц для каждого слоя требует использования трудоемких в вычислительном отношении математических функций $DSIN$ и $DCOS$, которые могут превосходить сложность операций, связанных с умножением матриц размером 2×2.

5 Заключение

В настоящее время на основе сочетания r-алгоритма с адаптивной регулировкой шага и случайного выбора начальной (стартовой) точки авторами разрабатываются алгоритмические процедуры для нахождения и анализа локальных экстремумов задач А и В. Делается это с целью проверить какой из критериев (гладкий или один из негладких) дает более правдоподобные результаты для близких к реальным тестовых примеров.

Изложенный в работе материал естественным образом приводит к дальнейшим исследованиям МОП по двум направлениям:

(1) построение новых математических моделей для различных задач нахождения оптимальных параметров плоских многослойных оптических структур (например, обобщение задач A и B на случай, когда угол падения не равен нулю, либо когда имеет место потеря электромагнитной волной энергии при прохождении через оптические среды и т.д.);

(2) разработка и выбор наиболее оптимальных (по тем или иным критериям) алгоритмических процедур для решения данных задач с учетом их специфики (как на основе субградиентных методов, в частности, методов с растяжением пространства, так и методов второго порядка, с различными адаптивными регулировками параметров этих методов).

Авторы надеются, что опыт применения математического аппарата гладкой и негладкой оптимизации для решения простейших задач A и B, описанных в данной работе, позволит более детально разобраться со свойствами многоэкстремальных задач, связанных с нахождением оптимальных параметров плоских многослойных оптических структур.

СПИСОК ЛИТЕРАТУРЫ

1. Furman Sh.A., Tikhonravov A.V. Basics of optics of multiplayer systems. – Editions Frontiers, Gif-sur Yvette, 1992. – 242 p.

2. Шор Н.З. Методы минимизации недифференцируемых функций и их приложения. – Киев: Наукова думка, 1979. – 199с.

3. Shor, N.Z. Nondifferentiable Optimization and Polynomial Problems, // Kluwer Academic Publishers, Boston / Dordrecht / London. – 1998. – 412 p.

4. Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. – М.: Мир, 1974, – 376с.

5. Нестеров Ю.Е. Эффективные методы в нелинейном программировании. – М.: Радио и связь, 1989. – 304 с.

6. Полак Э. Численные методы оптимизации. Единый подход. – М.: Мир, 1974. – 376 с.

7. Поляк Б.Т. Введение в оптимизацию. - М.: Наука, 1983. – 384с.

8. Пшеничный Б.Н., Данилин Ю.М. Численные методы в экстремальных задачах. – М.: Наука, 1975. – 319 с.

9. Химмельблау Д. Прикладное нелинейное программирование. – М.: Мир, 1975. – 535с.

10. О.В.Міца, П.І.Стецюк Задача знаходження оптимальних параметрів однорідного оптичного покриття // Теория оптимальных решений. – Киев: Ин-т кибернетики им. В.М.Глушкова НАН Украины, 2003. – С. 127-134