**XXXX期刊稿件编号**（由编辑填写）

# 非线性等式约束优化问题的正则化延拓法

**罗新龙*** · **肖航** · **张森**

arXiv:2303.14692v2 [math.OC] 4 Aug 2023

收稿日期：日期 / 接受日期：日期

**摘要** 本文研究了非线性等式约束优化问题的正则化延拓法和信赖域更新策略。具体而言，在适定阶段，采用正则化拟牛顿矩阵的逆作为预条件子，以提高计算效率；在病态阶段，则采用海森矩阵正则化双侧投影的逆作为预条件子，以增强算法的稳健性。由于该方法每次迭代只需求解一个线性方程组，而序列二次规划（SQP）每次迭代都需求解一个二次规划子问题，因此该方法比SQP更快。数值结果还表明，该方法比SQP更稳健、速度更快（SQP采用MATLAB2020a环境内置的子程序fmincon.m，以及在GAMS v28.2（2019年）环境中运行的子程序SNOPT）。对于大规模问题，新方法的计算时间约为fmincon.m的三分之一。最后，本文还给出了新方法的全局收敛性分析。

**关键词** 延拓法 ·预条件技术 ·信赖域方法 ·正则化技术 ·拟牛顿法

罗新龙（通讯作者）。中国北京市海淀区西土城路10号北京邮电大学人工智能学院，邮政信箱101，邮编100876。电子邮箱：luoxinlong@bupt.edu.cn 肖航 中国北京市海淀区西土城路10号北京邮电大学人工智能学院，邮政信箱101，邮编100876。电子邮箱：xiaohang0210@bupt.edu.cn 张森 中国北京市海淀区西土城路10号北京邮电大学人工智能学院，邮政信箱101，邮编100876。电子邮箱：senzhang@bupt.edu.cn

**数学主题分类（2010）**65K05 · 65L05 · 65L20 · 90C53

## 1 引言

本文考虑如下带有非线性等式约束的优化问题:

$$\begin{aligned} &\min_{x\in\Re^n} \ f(x) \\ &\text{subject to } \ c(x)=0, \end{aligned} \tag{1}$$

其中 $f\colon \Re^n \to \Re$ 和 $c\colon \Re^n \to \Re^m$ $(m \leq n)$。该问题在工程领域有许多应用，例如无人机保持水平飞行时的视觉惯性导航、 [11,49], 约束稀疏回归 [8], 稀疏信号恢复 [20,83], 图像复原和去噪 [21,57,77],Dantzig选择器 [43], 支持向量机 [22], 以及稀疏主成分分析（PCA）方法 [15,86,84]。此外，还有许多实用方法 [38,42,65] 以及许多高效的求解器（内置子程序 fmincon [72] MATLAB 2020a [60]中的子程序 SNOPT [29,30] GAMS 环境中的 [23]），可基于序列二次规划（SQP）方法求解该问题 [31,70,85]。

对于约束优化问题（1），延拓法 [3,12,24,37,67,80] 是传统优化方法（如序列二次规划 [31,70,71,85] 或罚函数法 [17]之外的另一种方法。与序列二次规划相比，延拓法的优点在于，它可以通过追踪非凸优化问题的轨迹找到许多局部最优点，甚至有可能找到全局最优解 [6,55,73,87]。然而，经典延拓方法的计算效率低于序列二次规划等传统优化方法。近年来，Luo 等人深入研究了求解非线性方程[50,53], 线性规划问题 [51], 线性互补问题 [56],无约束优化问题 [52,55],和线性约束优化问题 [49,54]的延拓方法，并提出了比传统优化方法更加稳健、快速的延拓方法。

在此，我们将他们的思路扩展到带非线性等式约束的优化问题。为了进一步提高延拓法求解大规模优化问题的计算效率，本文考虑将正则化BFGS（一种特殊的拟牛顿法）预条件技术用于正则化延拓法，并采用基于信赖域更新策略的自适应时间步长控制。此外，为了提高其稳健性，我们在病态阶段用海森矩阵正则化双侧投影的逆矩阵，替代正则化BFGS矩阵的逆矩阵。

本文其余部分安排如下。第 2 节针对问题（1）介绍采用切换预条件技术和信赖域更新策略的正则化延拓法。第 3 节分析该新方法的全局收敛性。第 4 节报告该方法的一些有前景的数值

结果，并将其与一些大规模问题上的最先进优化方法（SQP、MATLAB 2020a 内置的子程序 fmincon [72] 以及 [60] GAMS 环境中的子程序 SNOPT [29,30] [23]）进行比较。最后，我们在第 5 节进行讨论并给出结论。

## 2 正则化延拓法

本节介绍一种针对带非线性等式约束的优化问题的正则化延拓法，该方法结合了切换预条件技术和基于信赖域更新策略的自适应时间步进方案 [13] 。首先，我们考虑基于等式约束优化问题的 KKT 条件构造的正则化投影梯度流。随后，我们构造采用信赖域更新策略的正则化延拓法，以跟踪这类特殊的常微分方程（ODE）轨迹。新方法在适定阶段使用 Broyden-Fletcher-Goldfarb-Shanno（BFGS）更新公式 [5,18,25,76] 作为预条件技术，以提高计算效率；在病态阶段则采用正则化双侧投影 Hessian 矩阵的逆作为预条件子，以增强鲁棒性。此外，我们使用广义延拓牛顿法 [53] 来寻找初始可行点。

### 2.1 正则化投影牛顿流

对于非线性约束优化问题（1），众所周知，其最优解 $x^*$ 需要满足以下 Karush-Kuhn-Tucker 条件（第 328, [65] 页）：

$$\nabla_x L(x,\lambda) = \nabla f(x) + A(x)^T\lambda = 0, \tag{2}$$
$$c(x) = 0, \tag{3}$$

其中 $A(x)^T = [\nabla c_1(x)\ldots,\nabla c_m(x)]$ 且拉格朗日函数 $L(x\ \lambda)$ 定义为

$$L(x,\lambda) = f(x) + \lambda^T c(x). \tag{4}$$

与无约束优化问题中基于一阶必要条件（2)-（3）构造负梯度流的方法类似，我们可以为问题（1）构造一个微分代数方程动力系统， [33], [14,46,47,48, 49,54,74] 如下：

$$\frac{dx}{dt} = -\nabla L_x(x,\lambda) = -\left(\nabla f(x) + A(x)^T\lambda\right), \tag{5}$$
$$c(x) = 0. \tag{6}$$

对代数约束（6）关于$t$求导，并将其代入微分方程（5），得到

$$\begin{aligned}\frac{dc(x)}{dt} &= A(x)\frac{dx}{dt} = -A(x)\left(\nabla f(x) + A(x)^T\lambda\right)\\ &= -A(x)\nabla f(x) - A(x)A(x)^T\lambda = 0.\end{aligned} \tag{7}$$

如果进一步假设矩阵$A(x)$满行秩，则由方程（7）可得

$$\lambda = -\left(A(x)A(x)^T\right)^{-1}A(x)\nabla f(x). \tag{8}$$

将方程（$\lambda$8）代入方程（5），可得约束优化问题（[80] 1）的投影梯度流如下:

$$\frac{dx}{dt} = -\left(I - A(x)^T\left(A(x)A(x)^T\right)^{-1}A(x)\right)\nabla f(x) = -P(x)g(x), \tag{9}$$

其中，$g(x) = \nabla f(x)$，投影矩阵$P(x)$定义为

$$P(x) = I - A(x)^T\left(A(x)A(x)^T\right)^{-1}A(x). \tag{10}$$

不难验证$(P(x))^2 = P(x)$。也就是说，投影矩阵$P(x)$是对称的，其特征值为 0 或 1。根据[27],第 73 页中的定理 2.3.1 可知，其矩阵 2-范数为

$$\|P(x)\| = 1. \tag{11}$$

我们将$P(x)^\dagger$记作投影矩阵$P(x)$的摩尔-彭若斯广义逆（参见第 11, [79] 页或 [64,68]）。由于投影矩阵$P(x)$是对称的，且$(P(x))^2 = P(x)$，不难验证

$$P(x)^\dagger = P(x). \tag{12}$$

事实上，由方程 (12) 可知，$P(x)(P(x))^\dagger P(x) = P(x)$, $P(x)^\dagger P(x)P(x)^\dagger = P(x)^\dagger$, $\left(P(x)^\dagger P(x)\right)^T = P(x)^\dagger P(x)$ and $\left(P(x)P(x)^\dagger\right)^T = P(x)P(x)^\dagger$.

注 1 若 x(t) 是常微分方程 (9) 的解，不难验证 x(t) 满足 $A(x)(dx/dt) = 0$。也就是说，如果初始点 $x_0$ 满足 $c(x_0) = 0$，则投影梯度流 (9) 的解 x(t) 也满足 $c(x) = 0, \forall t \geq 0$。这一性质非常有用，因此我们可以构造一种常微分方程方法，沿着常微分方程 (9) 的轨迹前进，并求得其稳态解 $x^*$。□

假设$x(t)$是常微分方程（9）的解，利用性质$(P(x))^2 = P(x)$，可得

$$\begin{aligned}\frac{df(x)}{dt} &= (\nabla f(x))^T\frac{dx}{dt} = -(\nabla f(x))^T P(x)\nabla f(x)\\ &= -g(x)^T(P(x))^2 g(x) = -\|P(x)g(x)\|^2 \leq 0.\end{aligned}$$

也就是说，沿着动力系统（9）的解曲线 x(t)，f(x) 单调递减。此外，当 f(x) 有下界且 t 趋于无穷大时，解 x(t) 收敛到 $x^*$ [33,73,80], ，其中 $x^*$ 满足一阶 Karush-Kuhn-Tucker 条件（2）-（3）。因此，我们可以沿着常微分方程（9）的轨迹 x(t) 前进，得到其稳态解 $x^*$；该解也是原优化问题（1）的一个驻点。

然而，由于常微分方程 (9) 的右端函数 P(x)g(x) 秩亏，使用显式常微分方程方法跟踪投影梯度流 (9) [4,7,9] 时，我们将面临数值困难。为减轻常微分方程 (9) 的刚性，我们使用双侧投影矩阵 $P(x)\nabla^2 f(x)P(x)$ 的广义逆 $(P(x)\nabla^2 f(x)P(x))^\dagger$ 作为常微分方程 (9) 的预条件子，其用法与非线性方程组 [50], 无约束优化问题 [33,52,55], 线性规划问题 [51], 线性互补问题 [56], 欠定非线性方程组[49,54] 以及线性约束优化问题 [54]中的用法类似。

通过使用广义逆 $\left(P(x)\nabla^2 f(x)P(x)\right)^\dagger$ 作为 $P(x)\nabla^2 f(x)P(x)$ 的预条件子应用于常微分方程 (9)，我们得到投影牛顿流

$$\frac{dx(t)}{dt} = -\left(P(x)\nabla^2 f(x)P(x)\right)^\dagger P(x)g(x). \tag{13}$$

由于 $P(x)\nabla^2 f(x)P(x)$ 是奇异的，我们将方程（13）改写为

$$\left(P(x)\nabla^2 f(x)P(x)\right)\frac{dx(t)}{dt} = -P(x)g(x). \tag{14}$$

尽管投影牛顿流（14）减轻了常微分方程的刚性，使我们能够采用显式常微分方程方法在无穷区间上对其进行积分，但它仍有两个缺点。其一，双侧投影矩阵 $P(x)\nabla^2 f(x)P(x)$ 可能不是半正定的。因此，无法确保目标函数 f(x) 沿常微分方程（14）的解 x(t) 单调递减。其二，常微分方程（14）的解 x(t) 无法确保满足可行性，即 $dc(x)/dt = A(x)dx(t)/dt = 0$。为克服这两个缺点，我们对投影牛顿流（14）采用如下正则化技术 [32,81,82]：

$$\left(\sigma(x)I + P(x)\nabla^2 f(x)P(x)\right)\frac{dx(t)}{dt} = -P(x)g(x), \tag{15}$$

其中，正则化参数 $\sigma(x)$ 满足 $\sigma(x) + \mu_{min}\left(P(x)\nabla^2 f(x)P(x)\right) \geq \sigma_{min} > 0$。这里，$\mu_{min}(B)$ 表示矩阵 $B$ 的最小特征值。

备注 *2* 若假设 $x(t)$ 是常微分方程（15）的解，根据性质$A(x)P(x) = 0$，可得

$$A(x)\left(\sigma(x)I + P(x)\nabla^2 f(x)P(x)\right)\frac{dx(t)}{dt} = -A(x)P(x)g(x) = 0.$$

因此，我们得到 $\sigma(x)A(x)dx(t)/dt = 0$，即 $A(x)dx(t)/dt = 0$。对其积分可得 $c(x) = c(x_0) = 0$。也就是说，当 $c(x_0) = 0$ 时，常微分方程（15）的解 x(t) 保持可行性 $c(x) = 0$。□

注 3 由 $(P(x))^2 = P(x)$ 可知，常微分方程 (15) 的解 x(t) 满足 $P(x)dx(t)/dt = dx(t)/dt$。因此，由方程 (15) 和假设$\sigma(x) + \lambda_{min}\left(P(x)\nabla^2 f(x)P(x)\right) \geq \sigma_{min} > 0$，可得

$$\begin{aligned}\frac{df(x(t))}{dt} &= (\nabla f(x))^T \frac{dx(t)}{dt} = (\nabla f(x))^T P(x)\frac{dx(t)}{dt} = (P(x)g(x))^T \frac{dx(t)}{dt} \\ &= -(P(x)g(x))^T \left(\sigma(x)I + P(x)\nabla^2 f(x)P(x)\right)^{-1}(P(x)g(x)) \leq 0.\end{aligned}$$

也就是说，沿着常微分方程 (15) 的解 x(t)，f(x) 单调递减。此外，当 f (x) 有下界且 $\|P(x)\nabla^2 f(x)P(x)\| \leq M$ [33,41,73,80]，（其中 M 为正常数）时，解 x(t) 收敛到 $x^*$；$x^*$ 是正则化投影牛顿流 (15) 的驻点。因此，我们可以沿着常微分方程 (15) 的轨迹 x(t) 求得其驻点 $x^*$，该点也是原优化问题 (1) 的一个驻点。□

## 2.2 正则化延拓法

常微分方程（ODE）（$x(t)$的解曲线无法通过通用常微分方程方法高效跟踪，例如后向微分公式（BDF），即 MATLAB R2020a 环境中的子程序 ode15s.m [4,7,9,36] 。因此，我们需要为这个问题构造专门的方法。我们将一阶显式欧拉方法应用于常微分方程（ [75]15），由此得到如下正则化投影牛顿法:

$$\left(\sigma_k I + P(x_k)\nabla^2 f(x_k)P(x_k)\right)d_k = -P(x_k)g(x_k), \tag{16}$$

$$x_{k+1} = x_k + \alpha_k d_k, \tag{17}$$

其中 $\alpha_k$ 是时间步长。当 $\alpha_k = 1$时，正则化投影牛顿法（16）-（17）与 Levenberg-Marquardt 方法相似 [40,45,62]。

由于正则化投影牛顿法 (16)-(17) 的时间步长 $\alpha_k$ 受数值稳定性 [75]的限制。也就是说，对于线性测试方程 $dx/dt = -\lambda x$，其时间步长 $\alpha_k$ 受稳定区域 $|1 - \alpha_k\lambda/(\sigma_k + \lambda)| \leq 1$ 的限制。为避免这一缺点，类似于处理非线性方程 [50,51,53], 、无约束优化问题 [52,54] 和线性约束优化问题 [54],的技术，我们在方程 (17) 中用 $\Delta t_k/(1 + \Delta t_k)$ 替换 $\alpha_k$，并令方程 (16) 中的 $\sigma_k = \sigma_0/\Delta t_k$。由此得到正则化延拓法:

$$\left(\frac{\sigma_0}{\Delta t_k}I + B_k\right)d_k = -P_k g_k,\ s_k^p = \frac{\Delta t_k}{1 + \Delta t_k}(P_k d_k), \tag{18}$$

$$x_{k+1}^p = x_k + s_k^p, \tag{19}$$

其中， $\Delta t_k$ 为时间步长，$P_k = P(x_k)$，$g_k = \nabla f(x_k)$ 和 $B_k = P(x_k)\nabla^2 f(x_k)P(x_k)$，或其拟牛顿近似。

记 $A_k = A(x_k)$。矩阵 $A_k A_k^T$ 可能是病态的。因此，对于大规模问题，Cholesky分解方法可能无法求解由方程(10)定义的投影矩阵。因此，我们使用QR分解（第 247-248 页，[27]）来求解，如下所示：

$$A_k^T = Q_k R_k, \tag{20}$$

其中 $Q_k \in \Re^{n\times m}$ 满足 $Q_k^T Q_k = I$ 且 $R_k \in \Re^{m\times m}$ 是一个上三角矩阵。因此，由方程 (10)定义的投影矩阵 $P_k$ 可简化如下：

$$P_k = I - Q_k Q_k^T. \tag{21}$$

根据方程 (18)-(19) 以及投影矩阵 $\mathrm{P(x}_k\mathrm{)}$ 的性质 $\mathrm{A(x}_k\mathrm{)P(x}_k\mathrm{)} = 0$，不难验证 $\mathrm{A(x}_k\mathrm{)s}^p\mathrm{k} = 0$。因此，当约束为线性约束，即 $\mathrm{c(x)} = \mathrm{Ax} - \mathrm{b} = 0$ 时，若 $\mathrm{Ax}_k - \mathrm{b} = 0$ [54]，则 $\mathrm{x}^p\mathrm{k+1}$ 也满足线性约束 $\mathrm{Ax}^p\mathrm{k+1} - \mathrm{b} = 0$。然而，对于非线性约束 $\mathrm{c(x)} = 0$，这一性质并不成立。换言之，当 $\mathrm{c(x}_k\mathrm{)} = 0$ 时，我们可能有 $\mathrm{c}\left(\mathrm{x}^p\mathrm{k+1}\right) = 0$ 不成立，其中 $\mathrm{x}^p\mathrm{k+1}$ 是方程 (18)-(19) 的解。因此，我们需要计算校正步 $\mathrm{s}_k^c$，使 $\mathrm{x}_{k+1} = \mathrm{x}^p\mathrm{k+1} + \mathrm{s}_k^c$ 回到约束 $\mathrm{c(x)} = 0$ 上。为节省计算时间，我们采用以下带最短增量 [53] 的广义牛顿迭代来实现这一目标：

$$\begin{aligned} &s_k^c = -(A_{k+1}^p)^{\dagger} c(x_{k+1}^p),\ (A_{k+1}^p)^{\dagger} = (A_{k+1}^p)^T \left(A_{k+1}^p (A_{k+1}^p)^T\right)^{-1}, \\ &s_k \triangleq s_k^p + s_k^c,\ x_{k+1} = x_{k+1}^p + s_k^c = x_k + s_k, \end{aligned} \tag{22}$$

其中，$\mathrm{A}^p\mathrm{k+1}$ 等于 $\mathrm{A(x}_k\mathrm{)}$ 或 $\mathrm{A(x}^p\mathrm{k+1)}$。我们首先尝试 $\mathrm{A}^p\mathrm{k+1} = \mathrm{A(x}_k\mathrm{)}$，并求解方程（22）。如果 $\|\mathrm{c(x}_{k+1}\mathrm{)}\|_\infty > \varepsilon_0$，则尝试 $\mathrm{A}^p\mathrm{k+1} = \mathrm{A(x}^p\mathrm{k+1)}$，并再次求解方程（22），其中$\varepsilon_0$ 是给定的小常数，且小于终止容差 $\varepsilon$，例如 $\varepsilon_0 = \frac{1}{10}\varepsilon$ 和 $\varepsilon = 10^{-6}$。

当 $\mathrm{A}^p\mathrm{k+1} = \mathrm{A(x}_k\mathrm{)} = \mathrm{A}_k$ 时，我们已得到其 QR分解（20），并令 $\mathrm{Q}^p\mathrm{k} = \mathrm{Q}_k$，$\mathrm{R}^p\mathrm{k} = \mathrm{R}_k$。如果 $\mathrm{A}^p\mathrm{k+1} = \mathrm{A(x}^p\mathrm{k+1)}$，我们也使用 QR分解（第 247-248, [27] 页）将其分解为 $(\mathrm{A}^p\mathrm{k+1})^T = \mathrm{Q}^p\mathrm{k}\,\mathrm{R}^p\mathrm{k}$，其中 $\mathrm{Q}^p\mathrm{k} \in \Re^{n\times m}$ 满足 $(\mathrm{Q}^p\mathrm{k})^T\mathrm{Q}^p\mathrm{k} = \mathrm{I}$，且 $\mathrm{R}^p\mathrm{k} \in \Re^{m\times m}$ 是一个上三角矩阵。然后，我们按如下方式求解线性方程组（22）：

$$(R_k^p)^T d_k^c = -c(x_{k+1}^p),\ s_k^c = Q_k^p d_k^c,\ x_{k+1} = x_{k+1}^p + s_k^c. \tag{23}$$

我们将二次模型$q_k(s)$定义如下：

$$q_k(s) = f(x_k) + s^T g_k + \frac{1}{2} s^T B_k s, \tag{24}$$

其中，$g_k = \nabla f(x_k)$ 和 $B_k = P(x_k)\nabla^2 f(x_k)P(x_k)$ 或其拟牛顿近似。与零空间法的信赖域子问题（第571页—574, [79]）类似，我们将预测下降量分解为两部分：

$$Hpred_k = q_k(0) - q_k(s_k^p),\ Vpred_k = q_k(s_k^p) - q_k(s_k^p + s_k^c). \tag{25}$$

因此，为了确保正则化延拓法（18）-（23）的全局收敛，所接受的预测步 $s^p_k$ 和校正步 $s_k^c$ 需要满足以下条件：

$$\|s_k^c\| \le \theta_1 \|s_k^p\|, \tag{26}$$

其中 $\theta_1$ 是一个较大的正数，例如 $10^6$。

注 4 正则化延拓法（18）-（22）的时间步长 $\Delta t_k$ 不受数值稳定性的限制。因此，在稳态阶段可以采用较大的时间步长 $\Delta t_k$，使正则化延拓法（18）-（22）在驻点 $x^*$ 附近模拟投影牛顿法，从而获得较快的收敛速度。最重要的是，新的步长 $\alpha_k = \Delta t_k/(\Delta t_k + 1)$ 有利于采用信赖域更新策略来调整时间步长 $\Delta t_k$。因此，正则化延拓法（18）-（22）在瞬态阶段能够准确跟踪正则化流（15）的轨迹，并在驻点 $x^*$附近实现快速收敛。□

## 2.3 自适应步长控制

另一个问题是如何在每次迭代中自适应地调整时间步长 $\Delta t_k$。鉴于信赖域方法具有鲁棒性且收敛速度快，我们借鉴了该方法的调整技术 [13,88]。根据正则化延拓法 (18)-(22) 的结构保持性质，通过计算方程 (23) 的校正步 $s_k^c$，预期 $x_{k+1}$ 能够保持可行性。也就是说，$x_{k+1}$ 满足 $c(x_{k+1}) \approx 0$。因此，我们使用目标函数 $f(x)$，而非非光滑罚函数 $f(x) + \sigma\|c(x)\|_1$，作为优值函数。与用于无约束优化问题 [34,45,52] 和线性约束优化问题 [49,54], 的常微分方程方法步长方案类似，我们还需要构造 $f(x)$ 在 $x_k$附近的局部近似模型。这里，我们采用方程 (24) 定义的二次函数作为其近似模型。

我们将实际下降量$Ared_k$ 与预测下降量 $Pred_k$之比定义如下：

$$Ared_k \triangleq f(x_k) - f(x_k + s_k),\ Pred_k \triangleq q_k(0) - q_k(s_k),\ \rho_k = \frac{Ared_k}{Pred_k}. \tag{27}$$

因此，当 $\rho_k \ge \eta_a$, $\|c(x_k + s_k)\|_\infty \le \varepsilon_0$，且预测下降量 $Pred_k = q_k(0) - q_k(s_k)$ 满足Armijo充分下降条件时，我们接受试探步 $s_k$，并令 $x_{k+1} = x_k + s_k$：

$$Pred_k \ge \eta_q \|s_k^p\| \|p_{g_k}\|, \tag{28}$$

其中，$p_{g_k} = P(x_k)g(x_k) = P(x_k)\nabla f(x_k)$，$s^p k$ 由方程（18）计算；$\eta_a$, $\eta_q$ 是较小的正数常量，例如 $\eta_a = \eta_q = 10^{-6}$； $\varepsilon_0$ 小于容差误差 $\varepsilon$，例如 $\varepsilon_0 = \frac{1}{10}\varepsilon$ 和 $\varepsilon = 10^{-6}$。否则，我们舍弃该步，并令 $x_{k+1} = x_k$。

因此，我们根据 $\Delta t_{k+1}$ 在每次迭代中增大或减小时间步长，其依据是 $\rho_k$ 由方程（27）定义的比值。具体的调整策略如下：

$$\Delta t_{k+1} = \begin{cases} \gamma_1 \Delta t_k, \text{ if } (\text{the trial step } s_k \text{ is accepted and } \rho_k \geq \eta_2), \\ \Delta t_k, \text{ else if } (\text{the trial step } s_k \text{ is accepted and } \eta_1 < \rho_k < \eta_2), \\ \gamma_2 \Delta t_k, \text{ others}, \end{cases} \tag{29}$$

其中，常数根据我们的数值实验选取为 $\eta_1 = 0.25$、 $\gamma_1 = 2$、 $\eta_2 = 0.75$、 $\gamma_2 = 0.5$。

注 5 这种基于信赖域更新策略的新时间步进方案，相较于传统的线搜索策略 [44]具有一些优点。若使用线搜索策略和阻尼投影牛顿法 (16)-(17) 跟踪投影牛顿流 (15) 的轨迹 x(t)，为了在稳态阶段实现快速收敛，每次迭代中都从 1 开始尝试阻尼投影牛顿法的时间步长 $\alpha_k$，并多次将其减半。由于二次模型 $f(x_k) + \nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k$ 在瞬态阶段可能无法很好地近似 $f(x_k + s_k)$，因此时间步长 $\alpha_k$ 会很小。结果，线搜索策略在瞬态阶段会耗费不必要的试探步。然而，基于信赖域策略 (27)-(29) 的时间步长选取方案可以克服这一缺点。□

## 2.4 自适应预条件技术

对于大规模问题，数值计算双侧投影矩阵 $P(x_k)\nabla^2 f(x_k)P(x_k)$ 需要耗费大量时间。为了克服这一缺点，在适定阶段，我们使用 Broyden-Fletcher-Goldfarb-Shanno (BFGS) 拟牛顿矩阵（见 [5,18,25,58,76] 或第 222-230, [65] 页）来近似正则化延拓法 (18)-(23) 的双侧投影矩阵 $P(x_k)\nabla^2 f(x_k)P(x_k)$，如下所示：

$$B_{k+1} = \begin{cases} B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}, \text{ if } y_k^T s_k > 0, \\ B_k, \text{ otherwise}, \end{cases} \tag{30}$$

其中 $y_k = P(x_{k+1})g(x_{k+1}) - P(x_k)g(x_k)$，$s_k = x_{k+1} - x_k$ 且 $B_0 = I$。

BFGS更新矩阵$B_k$ 具有一些良好的性质。例如，$B_{k+1}$ 在$B_k$ 为对称正定且 $B_{k+1}$ 按 BFGS公式（30）（第 199, [65]）更新时，为对称正定。对于大规模问题，求解大型线性方程组（18）会耗费大量时间。为了

节省求解线性方程的计算时间，我们利用以下 Sherman-Morrison-Woodbury 公式（第 $(\sigma_k I + B_k)^{-1}$ 正则化矩阵 $(\sigma_k I + B_k)$ 求其逆矩阵 17, [79]）：

$$(B + SV^T)^{-1} = B^{-1} - B^{-1}S(I + V^T B^{-1} S)^{-1} V^T B^{-1}, \tag{31}$$

其中 $B$ 是一个$n \times n$ 非奇异矩阵，$S$、$V$ 是两个$p \times n$ 矩阵。一般来说，$p$ 大于$n$。我们将矩阵$S_k$ 和$V_k$ 表示如下：

$$S_k = \begin{cases} \left[ S_{k-1}, \frac{y_k}{(s_k^T y_k)^{1/2}}, \frac{m_k}{(s_k^T m_k)^{1/2}} \right], \text{if } y_k^T s_k > 0, \\ [S_{k-1}, 0, 0], \text{otherwise} \end{cases} \tag{32}$$

$$V_k = \begin{cases} \left[ V_{k-1}, \frac{y_k}{(s_k^T y_k)^{1/2}}, -\frac{m_k}{(s_k^T m_k)^{1/2}} \right], \text{if } y_k^T s_k > 0, \\ [V_{k-1}, 0, 0], \text{otherwise}, \end{cases} \tag{33}$$

其中 $m_k = B_k s_k = s_k + S_{k-1}(V_{k-1}^T s_k)$。于是，由方程（32）-（33），BFGS 准牛顿公式（30）可写为

$$B_{k+1} = I + S_k V_k^T. \tag{34}$$

利用 Sherman-Morrison-Woodbury公式（31），由方程（34）可得

$$(\sigma_k I + B_k)^{-1} = \frac{1}{1+\sigma_k} \left( I - S_{k-1} \left( (1+\sigma_k) I + V_{k-1}^T S_{k-1} \right)^{-1} V_{k-1}^T \right). \tag{35}$$

注*6* 值得讨论的是，方程（31）中的矩阵$I + V^T B^{-1} S$是否非奇异。实际上，当矩阵$B$ 为对称正定矩阵，且矩阵$B + SV^T$正定时（也就是说，$B + SV^T$的所有特征值都大于 0），$I + V^T B^{-1} S$是非奇异的。

证明。由于 B 是对称正定矩阵，根据 Cholesky分解（第 163, [27] 页），可将其分解为 B $= LL^T$，其中 L 是非奇异矩阵。因此，我们有 B $+ SV^T = L(I + (L^{-1}S)(V^T L^{-T}))L^T$。结合 B $+ SV^T$ 的正定性可知，$I + (L^{-1}S)(V^T L^{-T})$ 是正定的。因此，$(L^{-1}S)(V^T L^{-T})$ 的所有特征值都大于 -1。此外，$(L^{-1}S)(V^T L^{-T})$ 的特征值除零特征值外，均与 $(V^T L^{-T})(L^{-1}S)$ 的特征值相同。因此，$(V^T L^{-T})(L^{-1}S)$ 的所有特征值都大于 -1。也就是说，$V^T B^{-1} S$ 的所有特征值都大于 -1。因此，$I + V^T B^{-1} S$ 是一个非奇异矩阵。□

根据我们的数值实验 [49,54], ，正则化延拓法 (18)-(23) 配合 BFGS 更新公式 (32)-(35) 对大多数问题都效果良好，且目标函数在适定阶段下降得非常快。然而，对于病态问题，正则化延拓法 (18)-(23) 配合 BFGS 更新公式 (32)-(35) 将趋近于驻点解

$x^*$ 在病态阶段非常缓慢。此外，有时它无法接近驻点解$x^*$ 。

为了提高正则化延拓法 (18)- (23) 的鲁棒性，我们在病态阶段设置 $B_k = P(x_k)\nabla^2 f(x_k)P(x_k)$。现在的问题是，如何自动识别病态阶段，并从 BFGS 更新公式 (30) 切换到 $B_k = P(x_k)\nabla^2 f(x_k)P(x_k)$。这里我们采用一个简单的判据：一旦存在时间步长 $\Delta t_K \leq 10^{-3}$，我们就认为正则化延拓法 (18)-(23) 处于病态阶段。

在病态阶段，如果每次迭代都更新投影海森矩阵 $P(x_k)\nabla^2 f(x_k)P(x_k)$，其计算时间会很长。为了节省海森矩阵计算 $\nabla^2 f(x_k)$ 的计算时间，当 $q_k(0) - q_k(s_k)$ 能很好地近似 $f(x_k) - f(x_k + s_k)$ 时，我们设置 $B_{k+1} = B_k$，其中近似模型 $q_k(s)$ 由方程 (24) 定义。否则，我们会在病态阶段更新 $B_{k+1} = P(x_{k+1})\nabla^2 f(x_{k+1})P(x_{k+1})$。在病态阶段，一种实用的更新策略如下

$$B_{k+1} = \begin{cases} B_k, \text{ if } |1-\rho_k| \leq \eta_1, \\ P(x_{k+1})\nabla^2 f(x_{k+1})P(x_{k+1}), \text{ otherwise}, \end{cases} \tag{36}$$

其中 $\rho_k$ 由方程（24）-（27）定义，且 $\eta_1 = 0.25$。

## 2.5 寻找初始可行点

对于基于常微分方程组（15）的正则化延拓法，一个重要问题是如何找到初始可行点$x_0$。也就是说，初始点$x_0$需要满足$c(x_0) = 0$。这里，我们使用广义延拓牛顿法 [53] 求解欠定方程组$c(x) = 0$，因为该方法具有鲁棒性和高效性。为方便起见，我们对广义延拓牛顿法作简要介绍。其详细描述和收敛性分析可参见文献 [53]。

对于欠定方程组$c(z) = 0$，我们构造广义牛顿流

$$\frac{dz(\tau)}{d\tau} = -A(z)^{\dagger} c(z),\ z(\tau_0) = z_0, \tag{37}$$

其中 $A(z)^{\dagger}$ 是雅可比矩阵 $A(z)$（第 11, [79] 页或第 290, [27] 页）。接着，我们构造广义延拓牛顿法 [53]，以跟踪广义牛顿流（37），并按如下方式获得其稳态解$z^*$：

$$\Delta z_j = -\frac{\Delta\tau_j}{1+\Delta\tau_j} A_j^{\dagger} c(z_j),\ z_{j+1} = z_j + \Delta z_j, \tag{38}$$

其中 $A_j$ 等于 $A(x_j)$，或者根据以下比率取其近似值

$$r_j = \frac{\|c(z_j)\| - \|c(z_j + \Delta z_j)\|}{\|c(z_j)\| - \|c(z_j) + A_j \Delta z_j\|} = \frac{\|c(z_j)\| - \|c(z_j + \Delta z_j)\|}{(\Delta \tau_j / (1 + \Delta \tau_j))\|c(z_j)\|}. \tag{39}$$

换言之，我们根据以下策略更新雅可比矩阵 $A^{j+1}$：

$$A_{j+1} = \begin{cases} A_j, \text{ if } |1 - r_j| \leq \eta_1, \\ A(z_{j+1}), \text{ otherwise.} \end{cases} \tag{40}$$

对于实际问题，$AjA_j^T$ 可能是病态的。因此，对于大规模问题，可能无法通过 Cholesky 分解求解线性方程组（38）。为了提高其稳健性，我们采用 QR分解求解如下：

$$A_j^T = U_j W_j,\ W_j^T \Delta z_j^m = -c(z_j),\ \Delta z_j^N = U_j \Delta z_j^m,\ \Delta z_j = \frac{\Delta \tau_j}{1 + \Delta \tau_j} \Delta z_j^N, \tag{41}$$

其中，$U_j \in \Re^{n \times m}$ 满足$U_j^T U^j = I$，且$W_j \in \Re^{m \times m}$ 是一个上三角矩阵。实际应用中，为了节省分解矩阵$Aj+1$的计算时间，当$A_j$表现良好时，即 $|1 - r_j| \leq \eta_1$，根据更新公式（40），我们在方程（$Wj+1 = Wj$ 和$Uj+1 = Uj$ 中设定41）。

广义延拓牛顿法（ $\Delta \tau_j$的自适应时间步长38）通过以下信赖域更新策略进行调整：

$$\Delta \tau_{j+1} = \begin{cases} \gamma_1 \Delta \tau_j, & \text{if } \left|1 - r_j\right| \leq \eta_1, \\ \Delta \tau_j, & \text{else if } \eta_1 < \left|1 - r_j\right| < \eta_2, \\ \gamma_2 \Delta \tau_j, & \text{others,} \end{cases} \tag{42}$$

其中，根据我们的数值实验，常数选取为 $\gamma_1 = 2$，$\gamma_2 = 0.5$，$\eta_1 = 0.25$，$\eta_2 = 0.75$。

对于实际问题，可能无法获得解析雅可比矩阵 $A(zj)$。因此，在实际应用中，我们用其差分近似替代雅可比矩阵 $A(zj)$，如下所示：

$$A(z_k) \approx \left[ \frac{c(z_j + \varepsilon e_1) - c(z_j)}{\varepsilon}, \ldots, \frac{c(z_j + \varepsilon e_n) - c(z_j)}{\varepsilon} \right], \tag{43}$$

其中，$e_i$ 的各元素均为零，只有$i$个元素为 1；参数 $\varepsilon$ 可根据我们的数值实验选取为 $10^{-6}$。

根据以上描述，我们在算法 $z^*$ 中给出了广义延续牛顿法的详细实现，该方法采用信赖域更新策略 (GCNMtr)，用于寻找 $c(z) = 0$ 的初始可行点 1。

结合前几节的讨论和算法1，我们在算法2中给出了带非线性等式约束的优化问题（1）的正则化延拓法及信赖域更新策略的详细实现。

**算法 1** 广义延拓牛顿方法和信赖域更

欠定方程组的更新策略（GCNMTr 方法） 输入：函数 c：$\Re^n \to \Re^m$, m $\leq$n，初始点 $z_0$ （可选）以及容差 $\varepsilon$（可选）。输出：非线性方程的近似解 $z^*$。1：当未提供 $z_0$ 或 $\varepsilon$时，设置默认值 $z_0 = \text{ones}(n, 1)$ 和 $\varepsilon = 10^{-7}$。2：初始化参数： $\eta_a = 10^{-6}$, $\eta_1 = 0.25$, $\gamma_1 = 2$, $\eta_2 = 0.75$, $\gamma_2 = 0.5$, maxit $= 400$。3：设置 $\Delta\tau_0 = 10^{-2}$、成功试探步标志 $= 1$、itc $= 0$、j $= 0$。4：计算 $c_j = c(z_j)$。5：设置 $r_{j-1} = 0$。6：当 (itc <maxit) 时执行 7：如果（成功试探步标志 == 1），则 8：设置 itc = itc + 1。9：计算 $\text{Res}_j = \|c_j\|_\infty$。10：如果 ($\text{Res}_j < \varepsilon$)，则 11：跳出；12：结束 if 13：如果 ($|1 - r_{j-1}| > \eta_1$)，则 14：根据方程（43）计算 $A_j = A(z_j)$。15：使用 QR分解 $[\bar{U}_j, \bar{W}_j] = \text{qr}(A_j^T)$，并设置 $U_j = \bar{U}_j(:, 1:m)$, $W_j = \bar{W}_j(1:m, 1:m)$。16：否则 17：设置 $U_j = U_{j-1}$, $W_j = W_{j-1}$。18：结束 if 19：通过求解 $W_j^T \Delta z_j^m = -c_j$ 和 $\Delta z_j^N = U_j \Delta z_j^m$，得到牛顿步 $\Delta z_j^N$。20：结束 if 21：设置 $\Delta z_j = \Delta\tau_j/(1 + \Delta\tau_j)\Delta z_j^N$, $z_{j+1} = z_j + \Delta z_j$。22：计算 $c_{j+1} = c(z_{j+1})$。23：如果 $\left(\|c_j\| < \|c_{j+1}\|\right)$，则 24：设置 $r_j = -1$；25：否则 26：根据方程（39）计算比率 $r_j$。27：结束 if 28：根据信赖域更新策略（42）调整时间步长 $\Delta\tau_{j+1}$。29：如果 ($r_j \geq \eta_a$)，则 30：接受试探点 $z_{j+1}$。设置成功试探步标志 $= 1$。31：否则 32：设置 $z_{j+1} = z_j$, $c_{j+1} = c_j$, $\Delta z_{j+1}^N = \Delta z_j^N$, 成功试探步标志 $= 0$。33：结束 if 34：设置 j $\longleftarrow$ j + 1。35：结束 while

- -

- -

## 3 算法分析

本节分析正则化延拓方法（18）-（23）以及非线性等式约束优化问题的自适应时间步长控制（即算法 2）的全局收敛性。与无约束优化问题的信赖域方法 [69]以及线性约束优化问题的延拓方法 [49], 的结果类似，我们有以下关于 $q_k(0) - q_k(s^p{}_k)$ 的下界估计。

**引理 1** 假设二次模型 $q_k(s)$ 由方程 (24) 定义，且 $d_k$ 是方程 (18)的解。此外，假设时间步长 $\Delta t_k$

**算法 2** 用于含有

非线性等式约束的优化问题的正则化延拓法（Rcm 方法）

**输入：** 目标函数 $f$： $\Re^n \to \Re$，等式约束 $c(x)=0$，$c$： $\Re^n \to \Re^m$，初始点 $z_0$（可选），终止参数 $\varepsilon$（可选）。

**输出：** 最优近似解 $x^*$。

若未提供 $z_0$ 或 $\varepsilon$，则设置 $z_0$ =ones$(n,1)$ 或 $\varepsilon = 10^{-6}$。令 $\varepsilon_0 = \frac{1}{10}\varepsilon$.

初始化参数： $\eta_a = 10^{-6}$, $\eta_m = 10^{-6}$, $\eta_1 = 0.25$, $\gamma_1 = 2$, $\eta_2 = 0.75$, $\gamma_2 = 0.5$, $\sigma_0 = 10^{-5}, \Delta t_K = 10^{-3}$, 最大迭代次数 = 300。设置 $\Delta t_0 = 10^{-2}$, 病态阶段标志 = 0, 成功试探步标志 = 1, $s_{-1}=0$, $y_{-1}=0$, $\rho_{-1}=0$, $B_0 = I$, $H_0 = I$, $Q^b_{-1} = I$, $R^b_{-1} = I$, $S_{-1}$ = zeros$(n,\ 1)$, $V_{-1}$ = zeros$(n,\ 1)$, 它 = 0。

使用广义延拓牛顿法（GCNMTr 方法，算法1）求得一个从 $z_0$ 中得到一个初始可行点，并将该初始可行点记为 $x_0$。

令 $k=0$。计算 $f_0 = f(x_0)$ 和 $g_0 = \nabla f(x_0)$。根据方程（$A_0 = A(x_0)$ 43）计算。

通过 QR分解 $[\bar{Q}0, \bar{R}_0] = \mathrm{qr}(A_0^T)$对 $A_0^T$ 进行因式分解。

令 $Q_0 = \bar{Q}_0(:,1{:}\,m)$, $R_0 = \bar{R}_0(1{:}\,m, 1{:}\,m)$。

计算投影梯度 $p_{g_0} = P_0 g_0$, 其中 $P_0$ 由方程（21）计算。

**当** $((\|p_{g_k}\|_\infty > \varepsilon)$ 且（itc $<$ max itc）) **时**

  itc =itc + 1;

  **如果** $\Delta t_k < \Delta t_K$ **那么**

    设置标志 illposed_phase = 1。

  **结束如果**

  **如果** (标记为病态阶段 == 0) **则**

    **如果** (标记为成功试探步 == 1)**则**

      **如果** $k > 0$ **则**

        求解方程 (32)-(33) 以得到矩阵 $S_{k-1}$ 和 $V_{k-1}$。

      **else**

        令 $S_{-1}$ = 为零$(n,1)$, 并令 $V_{-1}$ = 为零$(n,1)$。

      **结束如果**

      求解方程（18）和（35）以得到 $d_k = -\left(\left(p_{g_k} - S_{k-1}\left(\left((1+\sigma_0/\Delta t_k)I + V\,Tk_{-1}S_{k-1}\right)^{-1}(V_{k-1}^T p_{g_k})\right)\right)\right)/(1+\sigma_0/\Delta t_k)$.

    **结束如果**

  **else**

    **如果** $(|\rho_{k-1}-1| > \eta_1)$ **则**

      计算 $H_k = P_k \nabla^2 f(x_k) P_k$ 由方程（115）。

      令 $B_k = (\sigma_0/\Delta t_k)I + H_k$，并使用 QR分解 $B_k = Q_k^b R_k^b$。

    **else**

      $Q_k^b = Q_{k-1}^b$, $R_k^b = R_{k-1}^b$;

    **结束条件判断**

    求解线性方程组 $R_k^b d_k = -(Q_k^b)^T p_{g_k}$ 以得到 $d_k$。

  **结束条件判断**

  设置 $s_k^p = \frac{\Delta t_k}{1+\Delta tk}(P_k d_k)$ 和 $x^p{}_{k+1} = x_k + s^p{}_k$。计算 $c^p{}_{k+1} = c(x^p_{k+1})$.

  求解线性方程组 $(R_k)^T d_k^c = -c^p{}_{k+1}$，$s_k^c = Q_k d_k^c$。

  设置 $x_{k+1} = x^p_{k+1} + 秒_k^c$，$c_{k+1} = c(x_{k+1})$。

  **如果** $(\|c_{k+1}\|_\infty > \varepsilon_0)$ **则**

    计算 $A^p{}_{k+1} = A(x^p_{k+1})$ 根据方程（43）。

    求解线性方程组（23），得到 $x_{k+1}$ 并计算 $c_{k+1} = c(x_{k+1})$。

  **结束 if**

  计算 $f_{k+1} = f(x_{k+1})$ 并根据方程（ $\rho_k$ 24）-（27）。

  **如果** ($\rho_k \geq \eta_a$, $\|c_{k+1}\|_\infty \leq \varepsilon_0$ 和 $s_k$ 满足充分下降条件（28）) **则**

    将成功试探步标志设为 = 1。计算 $g_{k+1} = \nabla f(x_{k+1})$。

    根据方程（43）计算$A_{k+1}$ =$A(x_{k+1})$ from equation (43).

    将 $A_{k+1}^T$ 按 $Q_{k+1}, \bar{R}_{k+1}] = \mathrm{qr}\,(A_{k+1}^T)$进行QR分解。

    令 $Q_{k+1} = \bar{Q}_{k+1}(:,1:m)$, $R_{k+1} = \bar{R}_{k+1}(1:m, 1:m)$。

    令 $p_{g_{k+1}} = P_{k+1} g_{k+1}$，其中 $P_{k+1}$ 由方程（21）计算得到。

    令 $s_k = x_{k+1} - x_k$ 和 $y_k = p_{g_{k+1}} - p_{g_k}$。

  **else**

    设置标志：成功试探步 = 0。

    $x_{k+1} = x_k$, $f_{k+1} = f_k$, $p_{g_{k+1}} = p_{g_k}$, $g_{k+1} = g_k$, $H_{k+1} = H_k$, $d_{k+1} = d_k$.

  **结束 if**

  根据信赖域更新策略调整时间步长 $\Delta t_{k+1}$ （29）。

  设定 $k \leftarrow k+1$。

**结束 while**

满足

$$\left(\frac{\sigma_0}{\Delta t_k} I + B_k\right) \succ 0 \; and \; \left(\frac{\sigma_0}{\Delta t_k} I + B_k - P_k^T B_k P_k\right) \succeq 0, \tag{44}$$

其中 $P_k = P(x_k)$ 是由方程(10)定义的投影矩阵。于是，我们有

$$q_k(0) - q_k(P_k d_k) \geq \frac{1}{2} \left\| p_{g_k} \right\| \min\left\{ \left\| P_k d_k \right\|, \frac{\left\| p_{g_k} \right\|}{3\|B_k\|} \right\}, \tag{45}$$

其中 $p_{g_k} = P_k g_k$。

**证明**。令 $\tau_k = \sigma_0 / \Delta t_k$。由方程（18），我们得到

$$\begin{aligned} q_k(0) - q_k(P_k d_k) &= -\frac{1}{2}\left(d_k^T P_k^T B_k P_k d_k\right) - \left(P_k g_k\right)^T d_k \\ &= -\frac{1}{2}\left(d_k^T P_k^T B_k P_k d_k\right) + p_{g_k}^T \left(\tau_k I + B_k\right)^{-1} p_{g_k} \\ &= \frac{1}{2}\left(p_{g_k}^T \left(\tau_k I + B_k\right)^{-1} p_{g_k} + d_k^T \left(-P_k^T B_k P_k + \tau_k I + B_k\right) d_k\right). \end{aligned} \tag{46}$$

我们将 $\mu_{\min}\left(B_k - P_k^T B_k P_k\right)$ 记为矩阵 $\left(B_k - P_k^T B_k P_k\right)$的最小特征值，并令

$$\tau_{lb} = \min\left\{0, \; \mu_{\min}\left(B_k - P_k^T B_k P_k\right)\right\}. \tag{47}$$

由方程（44）、（46）-（47）以及矩阵 $\left(\tau_k I + B_k\right)^{-1}$ 的特征值界，我们得到

$$\begin{aligned} q_k(0) - q_k(P_k d_k) &\geq \frac{1}{2}\left(p_{g_k}^T \left(\tau_k I + B_k\right)^{-1} p_{g_k} + \left(\tau_k + \tau_{lb}\right) \|d_k\|^2\right) \\ &\geq \frac{1}{2}\left(\frac{\left\| p_{g_k} \right\|^2}{\tau_k + \|B_k\|} + \left(\tau_k + \tau_{lb}\right) \|d_k\|^2\right). \end{aligned} \tag{48}$$

在上述第二个不等式中，我们使用性质 $|\mu_i(B_k)| \leq \|B_k\|$，其中 $\mu_i(B_k)$ 是矩阵$B_k$ 的一个特征值。

现在我们考虑该函数的性质

$$\varphi(\tau) \triangleq \tau \|d_k\|^2 + \left\| p_{g_k} \right\|^2 / \left(\tau - \tau_{lb} + \|B_k\|\right). \tag{49}$$

由方程（49），当（$\tau - \tau_{lb} + \|B_k\|$）$> 0$时，我们有 $\varphi''(\tau) = 2\|p_{g_k}\|^2 / (\tau - \tau_{lb} + \|B_k\|)^3 \geq 0$。因此，当 $\tau_{\min}$ 满足 $\varphi'(\tau_{\min}) = 0$ 且 $\tau \geq -(-\tau_{lb} + \|B_k\|)$ 时，函数 $\varphi(\tau)$ 取得其最小值 $\varphi(\tau_{\min})$。也就是说，我们有

$$\varphi(\tau_{\min}) = 2\|p_{g_k}\| \|d_k\| + (\tau_{lb} - \|B_k\|) \|d_k\|^2, \tag{50}$$

其中

$$\tau_{\min} = \frac{\left\| p_{g_k} \right\|}{\|d_k\|} + \tau_{lb} - \|B_k\|. \tag{51}$$

我们通过分别考察（45）来证明该性质： $\tau_{\min} \geq 0$ 或 $\tau_{\min} < 0$ ，如下所示。

(i) 当 $\left(\|p_{g_k}\|/\|d_k\| + (\tau_{lb} - \|B_k\|)\right) \geq 0$时，由方程 (51)，可得 $\tau_{\min} \geq 0$。根据假设 (44) 以及 $\tau_{lb}$ 的定义 ( $\tau_k \geq -\tau_{lb}$，可得 。因此，由方程 (48)–(51)，得到

$$q_k(0) - q_k(P_k d_k) \geq \frac{1}{2}\left((\tau_k + \tau_{lb})\|d_k\|^2 + \frac{\|p_{g_k}\|^2}{\tau_k + \|B_k\|}\right) = \frac{1}{2}\varphi(\tau_k + \tau_{lb}) \geq \frac{1}{2}\varphi(\tau_{\min})$$
$$= \frac{1}{2}\left(\|p_{g_k}\|\|d_k\| + \left(\|p_{g_k}\|\|d_k\| + (\tau_{lb} - \|B_k\|)\|d_k\|^2\right)\right) \geq \frac{1}{2}\|p_{g_k}\|\|d_k\|. \tag{52}$$

(ii) 另一种情况是 $\left(\|p_{g_k}\|/\|d_k\| + (\tau_{lb} - \|B_k\|)\right) < 0$。在这种情况下，由方程 (51)，可得 $\tau_{\min} < 0$。不难验证，当 $\varphi(\tau)$ 时，单调递增；且 $\tau \geq 0$ 和 $\tau_{\min} < 0$。根据 $\tau_{lb}$ 的定义 (47) 及性质 (11)，可得

$$|\tau_{lb}| \leq \left|\mu_{\min}\left(B_k - P_k^T B_k P_k\right)\right| \leq \|B_k - P_k^T B_k P_k\| \leq \|B_k\| + \|P_k^T B_k P_k\|$$
$$\leq \|B_k\| + \|P_k^T\|\|B_k\|\|P_k\| = 2\|B_k\|.$$

利用这一性质以及 $\varphi(\tau)$的单调性，由方程（48）—（49）可得

$$q_k(0) - q_k(P_k d_k) \geq \frac{1}{2}\left((\tau_k + \tau_{lb})\|d_k\|^2 + \frac{\|p_{g_k}\|^2}{\tau_k + \|B_k\|}\right)$$
$$= \frac{1}{2}\varphi(\tau_k + \tau_{lb}) \geq \frac{1}{2}\varphi(0) = \frac{1}{2(-\tau_{lb} + \|B_k\|)}\|p_{g_k}\|^2 \geq \frac{1}{6\|B_k\|}\|p_{g_k}\|^2. \tag{53}$$

由方程（52）—（53）可得

$$q_k(0) - q_k(P_k d_k) \geq \frac{1}{2}\|p_{g_k}\| \min\left\{\|d_k\|, \frac{\|p_{g_k}\|}{3\|B_k\|}\right\}. \tag{54}$$

利用矩阵$P_k$的性质（11），可得

$$\|P_k d_k\| \leq \|P_k\|\|d_k\| = \|d_k\|. \tag{55}$$

因此，由不等式（54）—（55）可得到估计式（45）。 □

当 $\Delta t^k \leq \frac{\sigma_0}{2\|B_k\|}$时，利用投影矩阵 $P_k$的性质 $\|P_k\| = 1$，可得

$$\left(\frac{\sigma_0}{\Delta t_k} I + B_k\right) \succ 0 \text{ and } \left(\frac{\sigma_0}{\Delta t_k} I + B_k - P_k^T B_k P_k\right) \succeq 0. \tag{56}$$

因此，由方程（18）和引理 1，当 $\Delta t^k \leq \frac{\sigma_0}{4\|B_k\|}$时，可得水平步预测下降量的估计如下。

**引理 2** 假设二次模型 $q_k(s)$ 由方程 (24) 定义，且 $s^p k$ 由正则化方法 (18)求得。那么，当 $\Delta t^k \leq {}_{4\|}{}^{\sigma_0} B_k\|$ 时，$q_0 - q_k(s^p k)$ 的下界估计如下： ——

$$Hpred_k = q_k(0) - q_k(s_k^p) = q_k(0) - q_k\left(\frac{\Delta t_k}{1+\Delta t_k} P_k d_k\right) \geq \frac{1}{2}\|P_k g_k\|\|s_k^p\|. \tag{57}$$

证明。我们通过考虑两种不同的情况来证明这一结果，即 $(s^p k)^T B_k s^p k \leq 0$ 或 $(s^p k)^T B_k s^p k > 0$。(i) 当 $(s^p k)^T B_k s^p k \leq 0$ 时，由方程（24）和方程（18）可得

$$\begin{aligned} q_k(0) - q_k(s_k^p) &= -g_k^T s_k^p - \frac{1}{2}(s_k^p)^T B_k s_k^p \geq -g_k^T s_k^p = -\frac{\Delta t_k}{1+\Delta t_k}(P_k g_k)^T d_k \\ &= \frac{\Delta t_k}{1+\Delta t_k}(P_k g_k)^T \left(\frac{\sigma_0}{\Delta t_k} I + B_k\right)^{-1} (P_k g_k) \geq \frac{\Delta t_k}{1+\Delta t_k} \frac{1}{\sigma_0/\Delta t_k + \|B_k\|}\|P_k g_k\|^2. \end{aligned} \tag{58}$$

此外，由方程（18）可知，当 $\Delta t^k < {}_{\|}{}^{\sigma_0} B_k\|$ 时，有

$$\begin{aligned} \|P_k g_k\| &= \left\|\left(\frac{\sigma_0}{\Delta t_k} I + B_k\right) d_k\right\| \geq (\sigma_0/\Delta t_k - \|B_k\|)\|d_k\| \\ &\geq (\sigma_0/\Delta t_k - \|B_k\|)\|P_k d_k\|. \end{aligned} \tag{59}$$

将方程（59）代入方程（58），当 $\Delta t^k \leq {}_{4\|}{}^{\sigma_0} B_k\|$ 时，得到——

$$\begin{aligned} q_k(0) - q_k(s_k^p) &\geq \frac{\Delta t_k}{1+\Delta t_k} \frac{\sigma_0/\Delta t_k - \|B_k\|}{\sigma_0/\Delta t_k + \|B_k\|}\|P_k g_k\|\|P_k d_k\| \\ &\geq \frac{\sigma_0/\Delta t_k - \|B_k\|}{\sigma_0/\Delta t_k + \|B_k\|}\|P_k g_k\|\|s_k^p\| \geq \frac{4\|B_k\| - \|B_k\|}{4\|B_k\| + \|B_k\|}\|P_k g_k\|\|s_k^p\| = \frac{3}{5}\|P_k g_k\|\|s_k^p\|. \end{aligned} \tag{60}$$

因此，对于情形 $(s^p k)^T B_k s^p k \leq 0$，结果（57）成立。

(ii) 当 $(s^p k)^T B_k s^p k \geq 0$，即 $(P_k d_k)^T B_k (P_k d_k) \geq 0$，由方程（24）和方程（18），可得

$$\begin{aligned} q_k(0) - q_k(s_k^p) &= -g_k^T s_k^p - \frac{1}{2}(s_k^p)^T B_k s_k^p \\ &= -\frac{\Delta t_k}{1+\Delta t_k} g_k^T (P_k d_k) - \frac{1}{2}\left(\frac{\Delta t_k}{1+\Delta t_k}\right)^2 (P_k d_k)^T B_k (P_k d_k) \\ &\geq \frac{\Delta t_k}{1+\Delta t_k}\left(-g_k^T (P_k d_k) - \frac{1}{2}(P_k d_k)^T B_k (P_k d_k)\right) = \frac{\Delta t_k}{1+\Delta t_k}(q_k(0) - q_k(P_k d_k)). \end{aligned} \tag{61}$$

当 $\Delta t^k \leq {}_{4\|}{}^{\sigma_0} B_k\|$ 时，将方程（45）和方程（59）代入方程（61），得到

$$\begin{aligned} q_k(0) - q_k(s_k^p) &\geq \frac{1}{2}\frac{\Delta t_k}{1+\Delta t_k}\|P_k g_k\| \min\left\{\|P_k d_k\|, \frac{\|P_k g_k\|}{3\|B_k\|}\right\} \\ &\geq \frac{1}{2}\frac{\Delta t_k}{1+\Delta t_k}\|P_k g_k\| \min\left\{\|P_k d_k\|, \frac{(\sigma_0/\Delta t_k - \|B_k\|)\|P_k d_k\|}{3\|B_k\|}\right\} \\ &\geq \frac{1}{2}\frac{\Delta t_k}{1+\Delta t_k}\|P_k g_k\|\|P_k d_k\| = \frac{1}{2}\|P_k g_k\|\|s_k^p\|. \end{aligned} \tag{62}$$

因此，结果（57）对于情形（$s^p{}_k$）$^T B_k s^p{}_k > 0$也成立。 □

在接下来对算法 2 的分析中，$f(x)$ 和 $c(x)$ 假设满足假设1。

假设*1* 假设 $f(x)$ 二阶连续可微，且 $c(x)$ 连续可微；序列 $\{x_k\}$ 由算法 *2* 在开集 $S$ 中生成； $\nabla f(x)$ $\nabla^2 f(x)$, $A(x)$ 在 $S$ 上有上界。我们还假设$\|B_k\|(k = 0, 1, 2, \ldots)$ 一致有界。也就是说，存在三个正常数 $M_B$, $M_g$, $M_A$， 使得

$$\begin{aligned} &\left\|\nabla^2 f(x)\right\| \le M_B,\ \|A(x)\| \le M_A,\ \|A(x)^\dagger\| \le M_A, \forall x \in \mathrm{S}, \\ &\text{and } \|B_k\| \le M_B,\ \|\nabla f(x_k)\| \le M_g,\ k = 0, 1, 2, \ldots \end{aligned} \tag{63}$$

成立。

结合投影矩阵 $\|P_k\| = 1$ 的性质 $P_k$，根据假设（63），我们得到

$$\begin{aligned} &\left\|P_k \nabla^2 f(x_k) P_k\right\| \le \|P_k\| \left\|\nabla^2 f(x_k)\right\| \|P_k\| = \left\|\nabla^2 f(x_k)\right\| \le M_B, \\ &\|P_k B_k P_k\| \le \|P_k\| \|B_k\| \|P_k\| = \|B_k\| \le M_B,\ k = 0, 1, 2, \ldots. \end{aligned} \tag{64}$$

根据矩阵范数的性质，我们知道 $\mathrm{P_k \nabla^2 f(x_k) P_k}$ 的特征值绝对值小于 $\mathrm{M_B}$。若将 $\mu\left(\mathrm{P_k \nabla^2 f\,(x_k) P_k}\right)$ 记为 $\mathrm{P_k \nabla^2 f\,(x_k) P_k}$的特征值，则我们知道 $\left(\frac{\sigma_0}{\Delta}\mathrm{tk\ I} + \mathrm{P_k \nabla^2 f\,(x_k) P_k}\right)$ 的特征值为 $\frac{\sigma_0}{\Delta}\mathrm{tk} + \mu\left(\mathrm{P_k \nabla^2 (\,f_k) P_k}\right)$。因此，由等式（64）可知

$$\frac{\sigma_0}{\Delta t_k} I + P_k \nabla^2 f(x_k) P_k \succ 0, \text{ when } \Delta t_k < \frac{\sigma_0}{M_B}. \tag{65}$$

为了证明算法 2 的全局收敛性，我们需要证明迭代序列 $\{x_k\}$ 保持约束可行$c(x_k) \approx 0$，且其收敛点$x^*$ 满足 KKT条件 (2)。在下面的引理 3 中，我们证明由算法 2 生成的迭代序列 $\{x_k\}$ 在可行性$c(x) = 0$的邻域内保持可行，前提是时间步长 $\Delta t_k (k = 0, 1, \ldots)$ 足够小。收敛点$x^*$ 满足 KKT条件 (2) 的证明留待定理1给出。

**引理 3** 假设 $c(x)$ 连续可微，且其雅可比矩阵 $A(x)$是 *Lipschitz* 连续的。也就是说，存在一个正常数 $L_A$， 使得

$$\|A(x) - A(y)\| \le L_A \|x - y\|,\ \forall x, y \in \Re^n. \tag{66}$$

此外，假设 $s_k$ 由方程 (18)-(22) 求解得到，且 $B_k$, $\nabla f(x_k)(k = 0, 1, 2, \ldots)$ 一致有界。也就是说，它们满足假设 *1*。那么，存在一个正常数 $\delta_c$， 使得

$$\|c(x_k + s_k)\| \le \varepsilon_0, \tag{67}$$

当 $\Delta t_k \le \delta_c$， $0 < \varepsilon_0 \le$ 最小值$\{1/(L_A M_A^2)\ \ \varepsilon\}$ 且 $\|c(x_k)\| \le \varepsilon_0$。

**证明**。 当 $A^{p}{}_{k+1} = A(x_k)$ 且 $\|c(x^{k+1})\| \le \varepsilon_0$，我们得到结果（67）。因此，我们只需考虑以下情形：$A^{p}{}_{k+1} = A(x^{p}{}_{k+1})$ 根据一阶泰勒展开和方程（22），我们有

$$\begin{aligned} c(x_k+s_k) &= c(x_k+s_k^p+s_k^c) = c(x_{k+1}^p+s_k^c) = c(x_{k+1}^p) + \int_0^1 A(x_{k+1}^p+ts_k^c)s_k^c dt \\ &= c(x_{k+1}^p) + A(x_{k+1}^p)s_k^c + \int_0^1 (A(x_{k+1}^p+ts_k^c) - A(x_{k+1}^p))s_k^c dt \\ &= \int_0^1 (A(x_{k+1}^p+ts_k^c) - A(x_{k+1}^p))s_k^c dt. \end{aligned} \tag{68}$$

将A(x)的利普希茨连续性 (66)以及有界性假设$\|A(x)^{\dagger}\| \le M_A$ 代入方程 (68，可得

$$\begin{aligned} \|c(x_k+s_k)\| &\le \int_0^1 \|A(x_{k+1}^p+ts_k^c) - A(x_{k+1}^p)\|\|s_k^c\| dt \le \frac{1}{2}L_A\|s_k^c\|^2 \\ &= \frac{1}{2}L_A \left\|A(x_{k+1}^p)^{\dagger}c(x_{k+1}^p)\right\|^2 \le \frac{1}{2}L_A \left\|A(x_{k+1}^p)^{\dagger}\right\|^2 \left\|c(x_{k+1}^p)\right\|^2 \\ &\le \frac{1}{2}L_A M_A^2 \left\|c(x_{k+1}^p)\right\|^2. \end{aligned} \tag{69}$$

为了估计 $\|c(x_k+s_k)\|$ 的上界，我们需要估计 $\|c(x^{p}{}_{k+1})\|$ 的上界。根据投影矩阵 $P_k$ 的性质 $A_kP_k = 0$，有

$$A_k s_k^p = \frac{\Delta t_k}{1+\Delta t_k} A_k P_k d_k = 0. \tag{70}$$

因此，根据一阶泰勒展开、性质 (70) 以及 A(x)的利普希茨连续性 (66)，有

$$\begin{aligned} \|c(x_{k+1}^p)\| &= \left\|c(x_k) + \int_0^1 A(x_k+ts_k^p)s_k^p dt\right\| \\ &= \left\|c(x_k) + \int_0^1 (A(x_k+ts_k^p) - A(x_k))s_k^p dt\right\| \\ &\le \|c(x_k)\| + \int_0^1 \|A(x_k+ts_k^p) - A(x_k)\|\|s_k^p\| dt \le \|c(x_k)\| + \frac{1}{2}L_A\|s_k^p\|^2. \end{aligned} \tag{71}$$

此外，由方程（18）以及关于 $g(x)$和 $B_k$的假设（63），当 $\Delta t_k \le \sigma_0/(2M_B)$时，可得

$$\begin{aligned} \|s_k^p\| &= \frac{\Delta t_k}{1+\Delta t_k}\left\|\left(\frac{\sigma_0}{\Delta t_k}I + B_k\right)^{-1} P_k g_k\right\| \le \frac{\Delta t_k}{1+\Delta t_k}\frac{\|P_k g_k\|}{\sigma_0/\Delta t_k - \|B_k\|} \\ &\le \frac{\Delta t_k}{1+\Delta t_k}\frac{\|P_k\|\|g_k\|}{\sigma_0/\Delta t_k - M_B} \le \frac{\Delta t_k}{1+\Delta t_k}\frac{M_g}{2M_B - M_B} = \frac{\Delta t_k}{1+\Delta t_k}\frac{M_g}{M_B}. \end{aligned} \tag{72}$$

根据归纳假设，有 $\|c(x_k)\| \le \varepsilon_0$。将其代入方程（72）和方程（71），当 $\Delta t_k \le \sigma_0/(2M_B)$时，可得

$$\|c(x_{k+1}^p)\| \le \varepsilon_0 + \frac{1}{2}L_A\left(\frac{\Delta t_k}{1+\Delta t_k}\frac{M_g}{M_B}\right)^2. \tag{73}$$

因此，由方程（69）和方程（73），当 $\Delta t_k \le \sigma_0/(2M_B)$ 时，我们证明$\|c(x_k + s_k)\| \le \varepsilon_0$ 只需证明

$$\frac{1}{2}L_A M_A^2\left(\varepsilon_0 + \frac{1}{2}L_A\left(\frac{\Delta t_k}{1+\Delta t_k}\frac{M_g}{M_B}\right)^2\right)^2 \le \varepsilon_0.$$

也就是说，当 $\Delta t_k \le \sigma_0/(2M_B)$ 时，我们只需证明

$$\frac{\Delta t_k}{1+\Delta t_k} \le \frac{2M_B^2}{L_A M_g^2}\left(-\varepsilon_0 + \sqrt{\frac{2\varepsilon_0}{L_A M_A^2}}\right). \tag{74}$$

我们选取

$$\delta_c = \min\left\{\frac{\sigma_0}{4M_B}, \frac{\frac{2M_B^2}{L_A M_g^2}\left(-\varepsilon_0 + \sqrt{\frac{2\varepsilon_0}{L_A M_A^2}}\right)}{\left|1 - \frac{2M_B^2}{L_A M_g^2}\left(-\varepsilon_0 + \sqrt{\frac{2\varepsilon_0}{L_A M_A^2}}\right)\right|}\right\}. \tag{75}$$

然后，当 $\Delta t_k \le \delta_c$ 且 $\varepsilon_0 \le \min\{1/(L_A M_A^2)\ \varepsilon\}$ 时，我们知道方程（74）成立。也就是说，当 $\Delta t_k \le \delta_c$ 且 $\varepsilon_0 \le \min\{1/(L_A M_A^2)\ \varepsilon\}$ 时，有 $\|c(x_k + s_k)\| \le \varepsilon_0$。□

为了估计时间步数 $\Delta t_k (k = 1,\ 2,\ .....)$ 的下界，我们需要证明校正步 $s_k^c$ 小于预测步 $s^p{}_k$，并且预测垂直约化量 $Vpred_k$ 小于预测水平约化量 $Hpred_k$，其中$s_k^c$ 由方程（22）求得，$s^p{}_k$ 由方程（18）求得，而$Vpred_k$、$Hpred_k$由方程（25）定义。在一些假设下，它们的证明由下述引理 4 给出。

引理 *4* 假设 $c(x)$ 连续可微，且其雅可比矩阵 $A(x)$ 满足利普希茨连续性条件（66）。此外，假设 $s^p\ k$ 由方程（18）求解，且 $B_k (k = 0, 1, 2, ...)$ 一致有界。也就是说，$B_k (k = 0, 1, 2, ...)$ 满足方程（63）。假设 $\|P_k g_k\| \ge \varepsilon (k = 0, 1, 2, ...)$。那么，存在五个正数常量 $\theta_c$, $\theta_v$, $\delta_0$, $\delta_v$ 和 $\delta_{\varepsilon_0}$，使得

$$\|s_k^c\| \le \theta_c \|s_k^p\|^2 \ and\ |Vpred_k| \le \theta_v \|P_k g_k\| \|s_k^p\|, \tag{76}$$

当 $\delta_0 \le \Delta t_k \le \delta_v$，$0 < \varepsilon_0 \le \delta_{\varepsilon_0}$ 且 $\|c(x_k)\| \le \varepsilon_0$，其中 $0 < \theta_v < 1/2$，$s_k^c$ 由方程(22)求解，且 $Vpred_k$ 由方程(25)定义。

**证明**。 由方程（18），当 $\Delta t_k < \sigma_0/\|B_k\|$时，我们有

$$P_k d_k = -P_k\left(\frac{\sigma_0}{\Delta t_k} I + B_k\right)^{-1} P_k g_k.$$

将其与柯西-施瓦茨不等式 $|x^T y| \le \|x\|\,\|y\|$ 以及 $P_k^2 = P_k$结合起来，当 $\Delta t_k < \sigma_0/\|B_k\|$时，我们有

$$\begin{aligned}\|P_k g_k\|\|P_k d_k\| \ge |g_k^T P_k^2 d_k| &= \left|(P_k g_k)^T \left(\frac{\sigma_0}{\Delta t_k} I + B_k\right)^{-1} (P_k g_k)\right| \\ &\ge \frac{1}{\sigma_0/\Delta t_k + \|B_k\|}\|P_k g_k\|^2. \end{aligned} \tag{77}$$

也就是说，当 $\Delta t_k < \sigma_0/\|B_k\|$时，由方程（77）可得

$$\|s_k^p\| = \left\|\frac{\Delta t_k}{1+\Delta t_k} P_k d_k\right\| = \frac{\Delta t_k}{1+\Delta t_k}\|P_k d_k\| \ge \frac{\Delta t_k}{1+\Delta t_k}\frac{1}{\sigma_0/\Delta t_k + \|B_k\|}\|P_k g_k\|. \tag{78}$$

因此，根据假设 $\|B_k\| \le M_B$ 和 $\|P_k g_k\| \ge \varepsilon$，当 $\delta_0 \le \Delta t_k < \sigma_0/M_B$时，由方程（78）可得

$$\|s_k^p\| \ge \frac{\delta_0}{1+\delta_0}\frac{1}{\sigma_0/\delta_0 + M_B}\varepsilon = \frac{\delta_0 \varepsilon}{(1+\delta_0)(\sigma_0/\delta_0 + M_B)}. \tag{79}$$

由方程（22）、利普希茨连续性（66）和方程（71）可得

$$\begin{aligned}\|s_k^c\| &= \|A(x_{k+1}^p)^\dagger c(x_{k+1}^p)\| \le M_A\|c(x_{k+1}^p)\| = M_A\|c(x_k + s_k^p)\| \\ &\le M_A\left(\|c(x_k)\| + \frac{1}{2}L_A\|s_k^p\|^2\right) \le M_A\left(\varepsilon_0 + \frac{1}{2}L_A\|s_k^p\|^2\right). \end{aligned} \tag{80}$$

我们选取

$$\delta_{\varepsilon_0} = \min\left\{\left(\frac{\delta_0\varepsilon}{(1+\delta_0)(\sigma_0/\delta_0 + M_B)}\right)^2, \frac{1}{L_A M_A^2}\right\}. \tag{81}$$

然后，由方程（79）-（81），当 $\varepsilon_0 \le \delta_{\varepsilon_0}$时，可得

$$\|s_k^c\| \le M_A\left(1 + \frac{1}{2}L_A\right)\|s_k^p\|^2 = \theta_c\|s_k^p\|^2, \tag{82}$$

其中我们选取 $\theta_c = M_A(1+0.5L_A)$。由此，我们证明了方程（76）中的第一个不等式。

由方程（18），当 $\Delta t_k \le \sigma_0/(2M_B)$ 时，我们有

$$\begin{aligned}\|s_k^p\| = \left\|\frac{\Delta t_k}{1+\Delta t_k} P_k d_k\right\| \le \|P_k d_k\| \le \|d_k\| &= \left\|\left(\frac{\sigma_0}{\Delta t_k} I + B_k\right)^{-1} P_k g_k\right\| \\ \le \frac{1}{\sigma_0/\Delta t_k - M_B}\|P_k g_k\| &\le \frac{M_g}{M_B}. \end{aligned} \tag{83}$$

因此，由方程（25）、方程（82）和假设（63），我们有

$$|Vpred_k| = |q_k(s_k^p) - q_k(s_k^p + s_k^c)| = \left|(s_k^c)^T \nabla q_k(s_k^p) + \frac{1}{2}(s_k^c)^T \nabla^2 q_k(s_k^p) s_k^c\right|$$
$$= \left|(s_k^c)^T \left(g_k + B_k s_k^p + \frac{1}{2} B_k s_k^c\right)\right| \le \|s_k^c\| \left\|g_k + B_k s_k^p + \frac{1}{2} B_k s_k^c\right\|$$
$$\le \|s_k^c\|(\|g_k\| + \|B_k\|(\|s_k^p\| + \|s_k^c\|)) \le \theta_c(M_g + M_B(\|s_k^p\| + \|s_k^c\|))\|s_k^p\|^2. \quad (84)$$

将方程（82）和方程（83）代入方程（84），当 $\Delta t_k \le \sigma_0/(2M_B)$ 时，我们得到

$$|Vpred_k| \le \frac{\theta_c(2M_gM_B + \theta_c M_g^2)}{M_B}\|s_k^p\|^2 \le \frac{\theta_c(2M_gM_B + \theta_c M_g^2)}{M_B(\sigma_0/\Delta t_k - M_B)}\|P_k g_k\|\|s_k^p\|. \quad (85)$$

我们选取

$$\delta_v = \min\left\{\frac{\sigma_0}{4M_B}, \delta_c, \frac{\sigma_0 M_B \theta_v}{(2M_gM_B + \theta_c M_g^2)\theta_c + M_B^2\theta_v}\right\}, \quad (86)$$

其中，正数 $\delta_c$ 由方程（75）定义。因此，由方程（85）-（86），当 $\Delta t_k \le \delta_v$ 时，有 $|Vpred_k| \le \theta_v\|P_k g_k\|\|s^p{}_k\|$。我们选取足够小的正数 $\delta_0$ 使其满足 $\delta_0 \le \delta_v$。然后，结合方程（82），当 $\delta_0 \le \Delta t_k \le \delta_v$ 且 $\varepsilon_0 \le \delta_{\varepsilon_0}$ 时，我们得到结果（76）。□

注 7 我们选取

$$\theta_1 \ge \frac{\theta_c M_g}{M_B}. \quad (87)$$

然后，由方程（82）-（87），当 $\delta_0 \le \Delta t_k < \sigma_0/(2M_B)$ 且 $\varepsilon_0 \le \delta_{\varepsilon_0}$ 时，有

$$\|s_k^c\| \le \theta_1\|s_k^p\|. \quad (88)$$

也就是说，条件（26）成立。 □

为了证明当$k$ 趋于无穷时$pg_k$ 收敛到零，我们需要估计时间步长 $\Delta t_k (k = 1, 2, .....)$ 的下界。利用引理 2、引理 3 和引理 4，我们可以得到时间步长 $\Delta t_k$ ($k$ =1，2，.....) 的下界如下。

**引理 5** 假设 $f$： $\Re^n \to \Re$ 是二阶连续可微的，且 $\nabla f(x)$ 是 *Lipschitz* 连续的。也就是说，存在一个正常数 $L_g$ 使得

$$\|\nabla f(x) - \nabla f(y)\| \le L_g\|x - y\|, \forall x, y \in \Re^n. \quad (89)$$

假设 $c(x)$ 连续可微，且其雅可比矩阵 $A(x)$ 满足利普希茨连续性 (66)。假设序列 $\{x_k\}$ 由算法 *2* 生成，且假设条件 *1* 成立。此外，我们假设

$$\|P_k g_k\| > \varepsilon \quad (90)$$

对所有 $k=0,1,2,\ldots$, 其中 $g_k=g(x_k)=\nabla f(x_k), P_k=P(x_k)$ 且投影矩阵 $P(x)$由方程 (10)定义。因此，存在一个正常数 $\delta_{\Delta t}$使得

$$\Delta t_k \geq \gamma_2 \delta_{\Delta t} > 0,\ k=0,1,2,\ldots, \tag{91}$$

其中 $\Delta t_k$ 由信赖域更新策略自适应调整(27)-(29)。

**证明**。 根据引理 3，由归纳法可知，当 $\delta_0 \leq \Delta t_k \leq \delta_v$ 且 $0<\varepsilon_0 \leq \delta_{\varepsilon_0}$时，有

$$\|c(x_k)\| \leq \varepsilon_0,\ k=0,1,2,\ldots, \tag{92}$$

其中 $\delta_{\varepsilon_0}$ 由方程 (81定义，且 $\delta_v$ 由方程 (86定义。

由方程 (11) 可知 $\|P_k\|=1$。利用这一性质和假设 (63)，可得

$$\begin{aligned} &\left|\mu_{\min}\left(B_k-P_k^T B_k P_k\right)\right| \leq \left\|B_k-P_k^T B_k P_k\right\| \\ &\quad \leq \|B_k\|+\|P_k^T\|\|B_k\|\|P_k\| = 2\|B_k\| \leq 2M_B,\ k=0,1,2,\ldots, \end{aligned} \tag{93}$$

其中 $\mu_{\min}(B)$表示矩阵 $B$的最小特征值。因此，由方程 (93)可得

$$\begin{aligned} \mu_{\min}\left(\frac{\sigma_0}{\Delta t_k} I+B_k-P_k^T B_k P_k\right) &= \frac{\sigma_0}{\Delta t_k}+\mu_{\min}\left(B_k-P_k^T B_k P_k\right) \\ &\geq \frac{\sigma_0}{\Delta t_k}-2M_B,\ k=0,1,2,\ldots. \end{aligned} \tag{94}$$

同样，由假设（63），可得

$$\mu_{\min}\left(\frac{\sigma_0}{\Delta t_k} I+B_k\right) = \frac{\sigma_0}{\Delta t_k}+\mu_{\min}\left(B_k\right) \geq \frac{\sigma_0}{\Delta t_k}-M_B,\ k=0,1,2,\ldots. \tag{95}$$

因此，当 $\Delta t_k \leq \sigma_0/(2M_B)$ （$k=1,2,..$）时，方程（44）的正定条件满足。

根据$f$（$x$）的利普希茨连续性（89）以及柯西-施瓦茨不等式$|x^T y| \leq \|x\|\|y\|$，可得

$$\begin{aligned} &\left|\int_0^1 (\nabla f(x_k+ts_k)-\nabla f(x_k))^T s_k dt\right| \leq \int_0^1 \|\nabla f(x_k+ts_k)-\nabla f(x_k)\|\|s_k\| dt \\ &\leq \int_0^1 L_g\|s_k\|^2 t dt = \frac{1}{2}L_g\|s_k\|^2 = \frac{1}{2}L_g(\|s_k^p+s_k^c\|)^2 \leq \frac{1}{2}L_g(\|s_k^p\|+\|s_k^c\|)^2. \end{aligned} \tag{96}$$

将方程（76）和方程（83）代入方程（96），当 $\delta_0 \leq \Delta t_k \leq \delta_v$ 且 $0<\varepsilon_0 \leq \delta_{\varepsilon_0}$ 时，可得

$$\begin{aligned} &\left|\int_0^1 (\nabla f(x_k+ts_k)-\nabla f(x_k))^T s_k dt\right| \leq \frac{1}{2}L_g(\|s_k^p\|+\theta_c\|s_k^p\|^2)^2 \\ &\leq \frac{1}{2}L_g\left(1+\frac{\theta_c M_g}{M_B}\right)^2 \|s_k^p\|^2, \end{aligned} \tag{97}$$

and

$$\begin{aligned}&\frac{1}{2}|s_k^T B_k s_k| \le \frac{1}{2}\|B_k\|\|s_k\|^2 \le \frac{1}{2}M_B\|s_k^p + s_k^c\|^2 \le \frac{1}{2}M_B(\|s_k^p\| + \|s_k^c\|)^2\\&\le \frac{1}{2}M_B\left(\|s_k^p\| + \theta_c\|s_k^p\|^2\right)^2 = \frac{1}{2}M_B(1+\theta_c\|s_k^p\|)^2\|s_k^p\|^2\\&\le \frac{1}{2}M_B\left(1+\theta_c\frac{M_g}{M_B}\right)^2\|s_k^p\|^2. \end{aligned} \tag{98}$$

根据$f(x_k+s_k)$的一阶泰勒展开和方程（97），可得

$$\begin{aligned}&\left|f(x_k) - f(x_k+s_k) + \nabla f(x_k)^T s_k\right| = \left|\nabla f(x_k)^T s_k - \int_0^1 \nabla f(x_k+ts_k)^T s_k dt\right|\\&= \left|\int_0^1 (\nabla f(x_k) - \nabla f(x_k+ts_k))^T s_k dt\right| \le \frac{1}{2}L_g\left(1+\frac{\theta_c M_g}{M_B}\right)^2\|s_k^p\|^2. \end{aligned} \tag{99}$$

由方程（57）和方程（76），可得

$$\begin{aligned}&q_k(0) - q_k(s_k) = (q_k(0) - q_k(s_k^p)) - (q_k(s_k^p) - q_k(s_k^p + s_k^c))\\&\ge \frac{1}{2}\|P_k g_k\|\|s_k^p\| - \theta_v\|P_k g_k\|\|s_k^p\| = (0.5-\theta_v)\|P_k g_k\|\|s_k^p\|. \end{aligned} \tag{100}$$

因此，由方程（24）、（27）、（57）、（76）以及（98）-（100），当$\delta_0 \le \Delta t_k \le \delta_v$且$0 < \varepsilon_0 \le \delta_{\varepsilon_0}$时，可得

$$\begin{aligned}|\rho_k - 1| &= \left|\frac{(f(x_k) - f(x_k+s_k)) - (q_k(0) - q_k(s_k))}{q_k(0) - q_k(s_k)}\right|\\&\le \frac{0.5(L_g+M_B)(1+(\theta_c M_g)/M_B)^2\|s_k^p\|^2}{|q_k(0) - q_k(s_k)|}\\&\le \frac{0.5(L_g+M_B)(1+(\theta_c M_g)/M_B)^2\|s_k^p\|^2}{\big||q_k(0) - q_k(s_k^p)| - |q_k(s_k^p) - q_k(s_k^p+s_k^c)|\big|}\\&\le \frac{0.5(L_g+M_B)(1+(\theta_c M_g)/M_B)^2\|s_k^p\|^2}{(0.5-\theta_v)\|P_k g_k\|\|s_k^p\|}. \end{aligned} \tag{101}$$

将方程（83）代入方程（101），得到

$$|\rho_k - 1| \le \frac{0.5(L_g+M_B)(1+(\theta_c M_g)/M_B)^2}{(0.5-\theta_v)(\sigma_0/\Delta t_k - M_B)}. \tag{102}$$

我们记

$$\begin{aligned}&\delta_{ub} = \frac{(0.5-\theta_v)(1-\eta_2)\sigma_0}{0.5(L_g+M_B)\left(1+(\theta_c M_g)/M_B\right)^2 + (0.5-\theta_v)(1-\eta_2)M_B},\\&\delta_{\Delta t} \triangleq \min\left\{\delta_{ub}, \frac{\sigma_0}{4M_B}, \delta_v, \Delta t_0\right\}. \end{aligned} \tag{103}$$

因此，由方程 (102)-(103) 可知，当 $\delta_0 \le \Delta t_k \le \delta_{\Delta t}$ 时，有 $\rho_k \ge \eta_2$。我们选取 $\delta_0 \le \gamma_2 \delta_{\Delta t}$。假设 K 是满足 $\Delta t_K \le \delta_{\Delta t}$ 的第一个指标。那么，由方程 (102)-(103) 可知 $\rho_K \ge \eta_2$。根据时间步长调整公式 (29) 和引理 3，当 $0 < \varepsilon_0 \le \delta_{\varepsilon_0}$ 时，$x_K + s_K$ 将被接受，且时间步长 $\Delta t_{K+1}$ 将增大。因此，对所有 $k = 0, 1, 2, \ldots$，均有 $\Delta t_k \ge \gamma_2 \delta_{\Delta t}$。

□

注*8* 我们选择 $\eta_q \le 0.5 - \theta_v$。那么，当 $\delta_0 \le \Delta t_k \le \delta_v$ 时，由方程（100）可得

$$q_k(0) - q_k(s_k) \ge (0.5 - \theta_v)\|P_k g_k\|\|s_k^p\| \ge \eta_q \|P_k g_k\|\|s_k^p\|. \tag{104}$$

也就是说，满足充分下降条件（28）。 □

利用引理 3 和引理 5的结果，我们证明算法 2 对等式约束优化问题（1）的全局收敛性如下。

定理*1* 假设假设*1*成立，$\nabla f(x)$满足利普希茨连续性（*89*），且$c(x)$的雅可比矩阵$A(x)$满足利普希茨连续性（*66*）。设序列 $\{x_k\}$ 由算法 *2* 生成，且$f(x_k)$ $(k = 0, 1, \ldots)$有下界。那么，对于任意正数常量 $\varepsilon$，我们都可以选取一个充分小的常量 $\varepsilon_0 > 0$，使得

$$\|P_K g_K\| \le \varepsilon \ and \ \|c(x_K)\| \le \varepsilon_0 \tag{105}$$

当 $\|c(x_0)\| \le \varepsilon_0$时，对某个指标$K$成立。

**证明**。 我们用反证法证明这一结果。假设结果（105）不成立。那么，我们有

$$\|P_k g_k\| > \varepsilon \text{ for all } k = 0, 1, 2, \ldots. \tag{106}$$

因此，根据引理 3 和引理 5，我们可以选取一个充分小的常数 $\varepsilon_0 > 0$，并且存在一个无限子序列 $\{x_{k_i}\}$，使得试探步 $s_{k_i}$ 被接受，即 $\rho_{k_i} \ge \eta_a$，$i = 1, 2, \ldots$。否则，若从某个迭代索引开始所有步都被拒绝，则时间步长将持续减小，这与（91）矛盾。因此，由方程（27），我们有 $\|c(x^{k_i})\| \le \varepsilon_0$ 并且

$$f_0 - \lim_{k\to\infty} f_k = \sum_{k=0}^{\infty}(f_k - f_{k+1}) \ge \eta_a \sum_{i=0}^{\infty}\left(q_{k_i}(0) - q_{k_i}(s_{k_i})\right), \tag{107}$$

其中$s^{k_i}$ 由方程（18）和方程（22）计算得到。

由于 $f_k = f(x_k)(k = 0, 1, 2, \ldots\ldots)$ 有下界，且序列 $\{f_k\}$ 单调递减，因此我们知道序列 $\{f(x_k)\}$ 存在极限，记为 $\lim_{k\to\infty} f_k = f^*$。因此，由方程（107），可得

$$\lim_{k_i\to\infty}\left(q_{k_i}(0) - q_{k_i}(s_{k_i})\right) = 0. \tag{108}$$

将方程（28）代入方程（108），得到

$$\lim_{i\to\infty}\|p_{g_{k_i}}\|\|s_{k_i}^p\|=0. \tag{109}$$

根据假设 $\|P_k g_k\|>\varepsilon>0$，由方程（109），可得

$$\lim_{i\to\infty}\|s_{k_i}^p\|=0. \tag{110}$$

由引理 5可知，存在一个正常数 $\delta_{\Delta t}$ 使得

$$\Delta t_k\geq\gamma_2\delta_{\Delta t}>0,\ k=0,1,2,\ldots. \tag{111}$$

(i) 当 $B^{k_i}$ 由 BFGS 准牛顿公式（30）更新时，可知$B_{k_i}\succ 0$。因此，将方程（111）代入方程（78），可得

$$\begin{aligned}\|s_{k_i}^p\|&\geq\frac{\Delta t_{k_i}}{1+\Delta t_{k_i}}\frac{1}{\sigma_0/\Delta t_{k_i}+\|B_{k_i}\|}\|P_{k_i}g_{k_i}\|\\&\geq\frac{\gamma_2\delta_{\Delta t}}{1+\gamma_2\delta_{\Delta t}}\frac{1}{\sigma_0/(\gamma_2\delta_{\Delta t})+M_B}\|P_{k_i}g_{k_i}\|.\end{aligned} \tag{112}$$

(ii) 当 $B^{k_i}=P_{k_i}\nabla^2 f(x^{k_i})P^{k_i}$时，由方程（18），可得

$$\begin{aligned}&P_{k_i}\left(\frac{\sigma_0}{\Delta t_{k_i}}I+P_{k_i}\nabla^2 f(x_{k_i})P_{k_i}\right)d_{k_i}=-P_{k_i}^2 g_{k_i}\\&=-P_{k_i}g_{k_i}=\left(\frac{\sigma_0}{\Delta t_{k_i}}I+P_{k_i}\nabla^2 f(x_{k_i})P_{k_i}\right)d_{k_i},\end{aligned}$$

由此得到 $P^{k_i}d^{k_i}=d^{k_i}$。将其与方程（18）和方程（111）结合，可得

$$\begin{aligned}&\frac{\gamma_2\delta_{\Delta t}}{1+\gamma_2\delta_{\Delta t}}\|P_{k_i}g_{k_i}\|\leq\left\|\frac{\Delta t_{k_i}}{1+\Delta t_{k_i}}P_{k_i}g_{k_i}\right\|\\&=\left\|\frac{\Delta t_{k_i}}{1+\Delta t_{k_i}}\left(\frac{\sigma_0}{\Delta t_{k_i}}I+P_{k_i}\nabla^2 f(x_{k_i})P_{k_i}\right)d_{k_i}\right\|\\&=\left\|\frac{\Delta t_{k_i}}{1+\Delta t_{k_i}}\left(\frac{\sigma_0}{\Delta t_{k_i}}I+P_{k_i}\nabla^2 f(x_{k_i})P_{k_i}\right)P_{k_i}d_{k_i}\right\|\\&=\left\|\left(\frac{\sigma_0}{\Delta t_{k_i}}I+P_{k_i}\nabla^2 f(x_{k_i})P_{k_i}\right)s_{k_i}^p\right\|\\&\leq\left(\frac{\sigma_0}{\Delta t_{k_i}}+\|B_{k_i}\|\right)\|s_{k_i}^p\|\leq\left(\frac{\sigma_0}{\gamma_2\delta_{\Delta t}}+M_B\right)\|s_{k_i}^p\|.\end{aligned} \tag{113}$$

将方程（112）-（113）代入方程（110），可得

$$\lim_{i\to\infty}\|P_{k_i}g_{k_i}\|=0, \tag{114}$$

这与 $P_k g_k(k=0,1,2,\ldots)$ 的有界假设（106）相矛盾。因此，结果（105）成立。□

## 4 数值实验

本节进行一些数值实验，以测试算法 2 （Rcm 方法）的性能。代码在 MATLAB R2020a 环境中运行，使用的是一台配备英特尔四核中央处理器和 8GB 内存的惠普笔记本电脑 [60]。对于实际问题，通常无法获得解析海森矩阵 $\nabla^2 f(x_k)$。因此，在实践中，我们用以下差分近似替代海森矩阵的双侧投影 $P(x_k)\nabla^2 f(x_k)P(x_k)$:

$$P_k\nabla^2 f(x_k)P_k \approx \left[\frac{P_k g(x_k+\varepsilon P_k e_1)-P_k g(x_k)}{\varepsilon},\ldots,\frac{P_k g(x_k+\varepsilon P_k e_n)-P_k g(x_k)}{\varepsilon}\right], \tag{115}$$

其中，$e_i$ 的各元素均为零，只有第$i$个元素等于 1，$P_k$ =$P(x_k)$，参数 $\varepsilon$ 可根据我们的数值实验选为 $10^{-6}$ 。

序列二次规划 [19,25,65,85] 是约束优化问题的一种代表性方法。序列二次规划方法有两种典型的实现代码。一种是 MATLAB2020a 环境 [60,72]中的内置子程序 fmincon.m（NLPQL）。另一种是在 GAMS v28.2（2019年）环境 [23]中运行的子程序 SNOPT [29,30] 。fmincon 和 SNOPT 是求解非线性等式约束优化问题的两种稳健且高效的求解器。因此，我们选择这两种典型求解器作为比较基准。

我们从 CUTEst[26] 中选取 65 个非线性等式约束优化问题，并构造 60 个测试问题。因此，我们使用这 125 个问题来测试这三种算法（Rcm、fmincon 和 SNOPT）的性能。对于这 60 个构造的测试问题，我们使用阿克利函数 [78] 作为目标函数。阿克利函数可写为:

$$f(x) = -ae^{-b\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}} - e^{\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)} + a + e, \tag{116}$$

其中 $a$ $=20$，$b$ $=0.2$，$n$ $=2000$。然后，我们使用来自 [1,39,61,78] 的 20 个无约束优化问题的梯度作为其约束函数。为了测试该方法在不同约束维数下的性能，我们测试了约束较少的问题（$m$ $<<n$，例如 $m$ $=10$，$n$ $=2000$）、约束数中等的问题（$m$ $\approx \frac{1}{2}n$，例如 $m$ $=1000$，$n$ $=2000$）以及约束较多的问题（$m$ $\approx n$，例如 $m$ $=1999$，$n$ $=2000$）。

三种对比算法（Rcm、fmincon 和 SNOPT）的终止条件均由以下方式设定

$$\|\nabla_x L(x_k,\lambda_k)\|_\infty \le 1.0\times 10^{-6}, \tag{117}$$

$$\|c(x_k)\|_\infty \le 1.0\times 10^{-6},\ k=1,2,\ldots, \tag{118}$$

其中，拉格朗日函数 $L(x$ $\lambda)$ 由方程（4）定义，而 $\lambda$ 由方程（8）定义。我们对这 100 个问题进行测试，其中 $n$ $\approx 1000$ 到 $n$ $=2000$。数值结果列于表1-6中。65 个约束优化问题的数值结果

来自 CUTEst 的问题 [26] 列于表 1-3。此外，60 个构造的约束优化问题的数值结果列于表4-6。Rcm、fmincon 和 SNOPT 的计算时间及迭代次数分别见图 1-2。

由于算法 2 (Rcm) 使用算法 1 (GCNMTr) 来寻找初始可行点，我们在表 $1_G$中统计其迭代次数，其中包括 GCNMTr 的迭代次数（记为$itc_R$）以及 Rcm 自身的迭代次数（记为itc）1-6。SNOPT 给出求解二次规划子问题的次数（记为“Major”），并使用 SQOPT 求解器 [30] 求解二次规划子问题。SNOPT 还给出求解二次规划子问题的总迭代次数，记为“ICTS”。为确保评估的客观性，我们在表1-6中统计 SNOPT 求解二次规划子问题的次数及总迭代次数。fmincon 仅给出求解二次规划子问题的次数，我们在表 1-6中将其记为“步数”。

为了公平地评估和比较这三种方法（Rcm、fmincon 和 SNOPT），我们还采用性能剖面作为评估工具 [16]。求解器的性能剖面是某项性能指标的（累积）分布函数；这里将求解器的计算时间与所有求解器中的最佳计算时间之比作为性能指标。如果求解器无法求解某个问题，则将其比值设为较大的数值，例如 999。图 3 展示了三种非线性规划求解器（Rcm、fmincon 和 SNOPT）的性能剖面。我们还统计了 Rcm、fmincon 和 SNOPT 求解失败的问题数以及求解速度最快的问题数。统计结果列于表 7。

从表 1–6 和表 7 可以看出，Rcm 能够求解大多数测试问题，只有少数问题（6.4%）无法求解。然而，fmincon 无法求解约 28125 个 CUTEst 问题（22.4%），SNOPT 则无法求解约 21125 个问题（16.8%）。从表 1–6、图 2 和图 3 可以看出，对于大多数测试问题，Rcm 的计算时间显著短于 fmincon。从表 1–6 和图 1 还可以看出，对于大多数测试问题，Rcm 的迭代次数显著少于 SNOPT。原因之一是，Rcm 在每次迭代中只需求解一个维数为 n 的线性方程组。SQP 在每次迭代中求解二次规划子问题时，需要求解一个维数为 (m +n) 的线性方程组（第 531–532, [65] 页），并涉及约 23 $(m+n)^3$ 次浮点运算（第 116, [27] 页）。另一个原因是，Rcm 使用算法 1（GCNMTr）寻找初始可行点时，会采用约束 c(x) 的雅可比矩阵 A(x) 的自适应更新技术 $=0$。当 GCNMTr 求解欠定非线性方程组时，这项技术可以显著节省计算时间 [53]。我们还发现，在若干测试案例中，SNOPT 的速度显著快于 Rcm。这是因为编译型语言比解释型语言更具优势：SNOPT 使用 C++ 语言（编译型语言）编写，而 Rcm 使用 MATLAB 语言（解释型语言）编写。

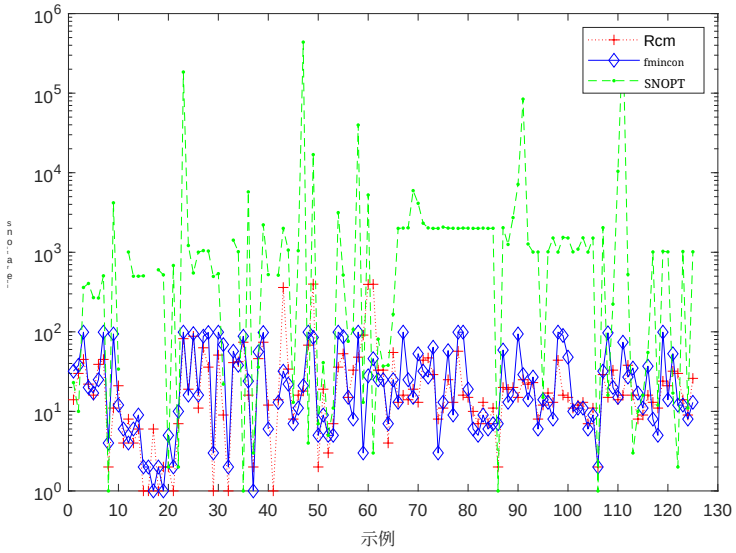



图 1：Rcm、fmincon 和 SNOPT 求解测试问题的迭代次数。

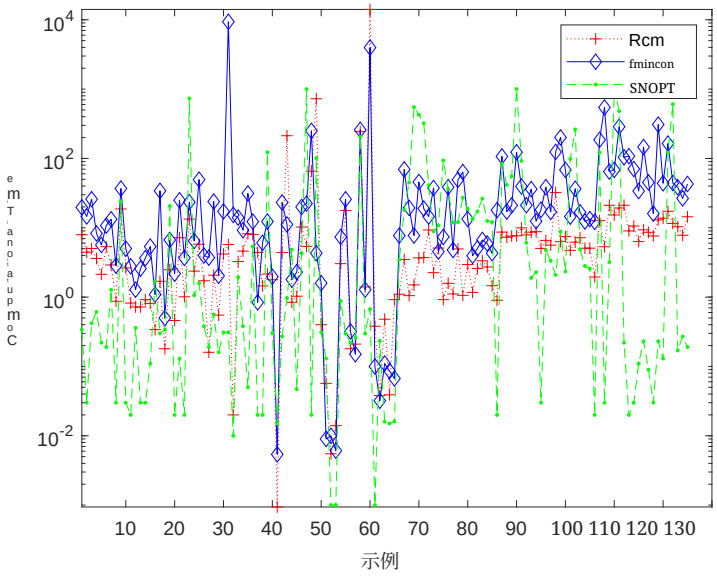



图 2：Rcm、fmincon 和 SNOPT 求解测试问题所用的 CPU 时间。

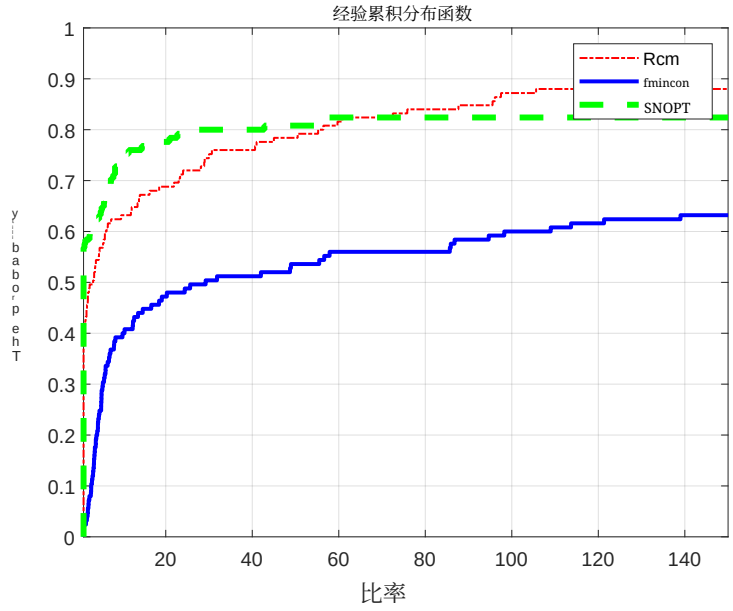



图 3：非线性规划求解器的性能剖面。

表 1：Rcm、fmincon 和 SNOPT 在 CUTEst 问题（例 1–20）上的数值结果。

| 问题 | Rcm | | fmincon | | SNOPT | |
|---|---|---|---|---|---|---|
| | $itc_G + itc_R$ (时间) | KKT ||约束（$x^{it}$）||$_\infty$ | 步数 (时间) | KKT ||约束（$x^{it}$）||$_\infty$ | 主要+ICTS (时间) | KKT ||约束($x^{it}$)||$_\infty$ |
| 考试. 1 LUKVLE1 [26] (n = 1000, m = 998) | 4+10 (7.93) | 4.59e-06 (2.66e-15) | 32 (19.68) | 1.32e-04 (1.78e-15) | 17+23 (0.34) | 1.40e-07 (8.55e-13) |
| 考试 2 LUKVLE2 [26] (n = 1000, m = 993) | 5+27 (4.43) | 3.96e-06 (2.50e-16) | 38 (14.57) | 2.67e-04 (5.82e-11) | 3+10 (0.03) | 1.70e+01 (2.51e+01) （失败） |
| 考试. 3 LUKVLE14 [26] (n = 1001, m = 666) | 4+41 (5.02) | 7.41e-06 (1.78e-15) | 99 (25.84) | 1.92e-03 (7.52e-08) (失败) | 25+361 (0.42) | 1.80e-06 (5.06e-13) |
| 考试. 4 LUKVLE11 [26] (n = 1001, m = 666) | 5+17 (3.61) | 4.41e-06 (8.88e-16) | 20 (8.29) | 4.00e-06 (8.88e-16) | 35+405 (0.61) | 1.20e-06 (3.84e-10) |
| 考试. 5 LUKVLE16 [26] (n = 1001, m = 750) | 3+13 (2.13) | 8.47e-06 (8.88e-16) | 17 (5.64) | 3.32e-06 (7.54e-12) | 18+269 (0.22) | 2.90e-07 (3.41e-13) |
| 例6 LUKVLE17 [26] (n = 1001, m = 750) | 4+35 (5.34) | 4.50e-06 (2.22e-16) | 25 (10.76) | 7.08e-06 (4.46e-11) | 15+266 (0.19) | 3.90e-07 (6.18e-09) |
| 例7 LUKVLE9 [26] (n = 1000, m = 6) | 1+50 (2.91) | 9.90e-06 (7.92e-07) | 99 (13.27) | 3.78 (8.20e-02) (失败) | 2+506 (1.27) | 4.00e+00 (3.10e-02) (失败) |
| 例 8 BDVALUE [26] (n = 1000, m = 1000) | 2+0 (0.87) | 0 (2.08e-08) | 4 (2.84) | 0 (4.83e-16) | 5+1 (0.03) | 9.10e-10 (1.72e-10) |
| 考试. 9 BDQRTIC [26] (n = 1000, m = 996) | 6+5 (18.68) | 2.81e-02 (1.10) （失败） | 94 (36.80) | 1.36 (3.30) （失败） | 222+4185 (23.86) | 1.60e-06 (1.62e+00) (失败) |
| 第10个测试 BDEXP [26] (n = 1000, m = 998) | 21+0 (2.63) | 0 (1.46e-07) | 12 (5.00) | 0 (9.53e-07) | 15+34 (0.03) | 1.80e-06 (3.10e-07) |
| 第11个测试 BROYDNBD [26] (n = 1000, m = 1000) | 4+0 (0.82) | 0 (1.47e-09) | 6 (2.78) | 0 (8.88e-16) | 7+0 (0.02) | 1.60e-13 (4.96e-10) |
| 考试. 12 YAO [26] (n = 1000, m = 1000) | 1+7 (0.72) | 1.33e-06 (0) | 4 (1.28) | 4.14e-08 (0) | 7+1007 (0.36) | 5.40e-07 (0) |
| 考试 13 WOODSNE [26] (n = 1000, m = 1000) | 4+0 (0.72) | 0 (7.23e-09) | 6 (2.57) | 0 (1.27e-12) | 7+500 (0.03) | 8.00e-07 (1.16e-07) |
| 考试 14 伍兹 [26] (n = 1000, m = 1000) | 6+0 (0.92) | 3.33e-16 (7.04e-08) | 9 (3.64) | 1.50e-08 (8.96e-14) | 10+500 (0.03) | 5.10e-10 (1.68e-08) |
| 考试. 15 SOSQP1 [26] (n = 1000, m = 501) | 1+0 (0.81) | 6.66e-15 (4.15e-10) | 2 (5.36) | 1.10e-13 (1.14e-13) | 3+505 (0.11) | 9.80e-12 (1.31e-11) |
| 考试 16 ARGLALE [26] (n = 1000, m = 1000) | 1+0 (0.34) | 0 (1.65e-08) | 2 (1.09) | 0 (3.29e-11) | 0+0 (1.69) | 0 (1.60e-12) |
| 考试 17 AUG2D [26] (n = 1000, m = 1000) | 1+0 (11.91) | 0 (0) | 1 (34.11) | 0 (0) | 0+0 (0.3) | 0.00e+0 (0) |
| 第18题 BLOCK二次规划1 [26] (n = 501, m = 1005) | 1+0 (0.18) | 5.08e-15 (9.99e-08) | 2 (0.50) | 1.55e-14 (4.44e-15) | 50+602 (0.34) | 2.70e-12 (1.65e+07) （失败） |
| 考试 19 BROWNAL [26] (n = 1000, m = 1000) | 2+0 (2.48) | 6.46e-14 (8.82e-08) | 1 (6.67) | 6.63e-06 (1.50e-06) | 22+520 (20.53) | 2.50e-08 (2.36e-11) |
| 考试. 20 BROYDEN3D [26] (n = 1000, m = 1000) | 2+0 (0.46) | 0 (3.72e-08) | 5 (2.18) | 5.08e-11 (1.11e-15) | 5+2 (0.02) | 1.70e-08 (1.56e-05) |

表 2：Rcm、fmincon 和 SNOPT 在 CUTEst 问题上的数值结果（考试 21-40）。

| 问题 | Rcm | | fmincon | | SNOPT | |
|---|---|---|---|---|---|---|
| | $itc_G+itc_R$ （时间） | KKT ||约束（$x^a$）||∞ | 步数 （时间） | KKT ||约束（$x^a$）||∞ | 主要+ICTS （时间） | KKT ||约束（$x^a$）||∞ |
| 考试 21 CHENHARK [26] （n = 1000， m = 1000） | 1+0 (7.18) | 9.76e-15 (8.25e-09) | 2 (24.94) | 1.55e-06 (1.55e-15) | 36+682 (0.13) | 3.00e-10 (4.75e-07) |
| 考试. 22 CHNROSNB [26] (n = 1000, m = 1000) | 7+0 (1.01) | 1.55e-15 (4.94e-08) | 10 (3.68) | 1.85e-08 (7.99e-15) | 11+2 (0.02) | 1.40e-08 (7.85e-08) |
| 示例 23 DIXON3DQ [26] (n = 1000, m = 0) | 0+83 (13.34) | 9.91e-06 (0) | 99 (23.44) | 6.66e-02 (0) （失败） | 168685+184416 (733.17) | 1.80e-06 (0) |
| 考试 24 DIXMAANA [26] （n = 1000， m = 0） | 0+16 (2.37) | 3.19e-06 (0) | 16 (6.45) | 3.81e-06 (0) | 12+1216 (1.06) | 2.60e-07 (0) |
| 考试 25 ORTHGDS [26] (n = 1003, m = 1000) | 1+40 (5.79) | 3.76e-08 (1.32e-07) | 95 (49.52) | 5.31e+02 (9.61e+02) （失败） | 41+549 (1.59) | 1.20e-06 (2.45e-11) |
| 考试 2 6 VARDIM [26] (n = 1000, m = 0) | 1+10 (1.72) | 1.60e-14 (0) | 16 (3.98) | 7.07e-06 (0) | 3+1003 (0.38) | 9.30e-06 (0) |
| 第27个算例 SVANBERG [26] (n = 1000, m = 1000) | 24+33 (0.16) | 6.04e-06 (6.34e-07) | 88 (3.71) | 8.64e+01 (2.12) （失败） | 0+1050 (0.19) | 2.00e+0 (2.10e+01) （失败） |
| 测试问题 28 SROSENBR [26] (n = 1000, m = 0) | 0+24 (2.07) | 1.88e-06 (0) | 98 (23.87) | 8.43e+02 (0) （失败） | 21+1036 (0.56) | 3.70e-08 (0) |
| 考试 2 9 SREADIN3 [26] (n = 1000, m = 1000) | 1+0 (0.55) | 2.08e-06 (2.08e-08) | 3 (2.03) | 2.08e-06 (3.06e-12) | 2+497 (0.16) | 1.90e-06 (7.43e-06) |
| Exam. 30 SOSQP2 [26] (n = 1000, m = 1000) | 1+50 (4.18) | 3.84e-06 (7.40e-07) | 99 (17.20) | 1.03e-04 (1.33e-11) | 38+538 (0.31) | 2.00e-06 (2.81e-11) |
| 例题 31 SINROSNB [26] (n = 1000, m = 1000) | 2+7 (5.73) | 5.17e-06 (2.22e-16) | 69 (9.36e+03) | 4.18e+04 (2.12e-02) （失败） | 18+22 (0.31) | 5.400e-07 (2.44e-13) |
| 考试. 32 SIPOW1 [26] (n = 1000, m = 2) | 1+0 (0.02) | 8.70e-19 (1.16e-08) | 2 (15.35) | 2.83e-06 (1.96e-18) | 0+0 (0.01) | 0 (0) |
| 考试. 33 SINQUAD [26] (n = 1000, m = 0) | 0+39 (3.26) | 8.65e-06 (0) | 57 (14.03) | 2.17e-06 (0) | 247+1419 (1.94) | 1.70e-06 (0) |
| 考试. 34 SINEALI [26] (n = 1000, m = 0) | 0+42 (4.61) | 8.80e-16 (0) | 38 (9.24) | 6.78e+03 (0) （失败） | 16+1017 (0.38) | 8.80e-07 (0) |
| 考试. 35 SEMICON1 [26] (n = 1000, m = 1000) | 74+0 (8.19) | 0 (1.63e-08) | 88 (31.28) | 0 (3.80e-13) | 9+1 (0.05) | 1.00e-01 (6.79e+02) （失败） |
| 考试。36 阅读1 [26] (n = 1000, m = 999) | 1+19 (7.98) | 4.26e-06 (5.28e-08) | 24 (12.27) | 1.25e-03 (1.04e-02) （失败） | 104+5740 (5.38) | 1.10e+03 (3.11e+05) （失败） |
| 考试 37 POWELL20 [26] (n = 1000, m = 1000) | 1+6 (4.41) | 3.10e-07 (1.14e-13) | 1 (0.84) | 3.46e-11 (3.67e-12) | 3+3 (0.02) | 2.27e-13 () |
| 第38题 LUKVLE3 [26] (n = 1000, m = 2) | 3+43 (1.45) | 9.52e-08 (5.98e-07) | 56 (6.07) | 2.67e-04 | 24+36 (0.02) | 1.00e-06 (3.52e-13) |
| 第39题 LUKVLE7 [26] (n = 1000, m = 4) | 4+54 (2.15) | 1.10e-06 (6.03e-07) | 98 (12.30) | 1.08e+05 (2.66e+03) （失败） | 1150+2210 (122.55) | 4.10e-04 (6.85e-12) |
| 考试. 40 LUKVLE12 [26] (n = 1001, m = 750) | 2+11 (1.78) | 7.24e-06 (8.88e-16) | 6 (1.97) | 1.63e-06 (3.41e-10) | 20+524 (0.30) | 1.70e-07 (2.62e-12) |

表3：Rcm、fmincon 和 SNOPT 在 CUTEst 问题上的数值结果（例41–65）。

| 问题 | Rcm | | fmincon | | SNOPT | |
|---|---|---|---|---|---|---|
| | $itc_G + itc_R$ (时间) | KKT $\|\|$约束$(x^{it})\|\|_\infty$ | 步数 (时间) | KKT $\|\|$约束$(x^{it})\|\|_\infty$ | 主要+ICTS (时间) | KKT $\|\|$约束$(x^{it})\|\|_\infty$ |
| 算例 41 飞机 [26] (n = 8, m = 5) | 1+0 (0.00094) | 0.00e+00 (0.00e+00) | 0 (0.0054) | 0.00e+00 (0.00e+00) | 0+0 (0.015) | 0.00e+00 (0.00e+00) |
| 考试。42 orthrdm2 [26] (n = 2003, m = 1000) | 4+10 (4.39) | 7.66E-06 (8.53E-14) | 13 (23.19) | 2.34E-06 (2.93E-10) | 6+509 (0.27) | 2.60E-07 (1.13E-06) |
| 考试 43 huestis [26] (n = 1000, m = 2) | 1+360 (212.67) | 472.58 (2.76E-08) (失败) | 32 (11.14) | 1.24E+03 (6.67E-06) (失败) | 5+1992 (0.97) | 3.70E-14 (9.09E-13) |
| 考试。44 吉尔伯特 [26] (n = 1000, m = 1) | 6+28 (0.84) | 8.58E-06 (2.24E-14) | 22 (1.79) | 8.58E-06 (9.34E-08) | 31+1035 (4.84) | 1.00E-06 (3.87E-09) |
| 考试。45 genhs28 [26] (n = 1000, m = 998) | 1+7 (1.03) | 3.32E-06 (2.22E-16) | 7 (2.30) | 1.51E-06 (2.22E-16) | 4+9 (0.047) | 3.60E-12 (4.55E-13) |
| 例46 dtoc6 [26] (n = 2001, m = 1000) | 3+13 (10.24) | 8.03E-06 (1.63E-13) | 11 (19.95) | 1.68E-06 (2.32E-12) | 20+1025 (4.27) | 1.70E-06 (2.91E-07) |
| 考试 47 dtoc5 [26] (n = 2001, m = 1000) | 1+17 (5.38) | 9.72E-06 (8.34E-12) | 21 (22.35) | 6.31E-06 (6.53E-09) | 155561+282454 (1000) | 4.90E-03 (9.54E-01) (失败) |
| Exam. 48 orthregc [26] (n = 2005, m = 1000) | 8+60 (65.61) | 8.76E-06 (8.07E-07) | 99 (250.32) | 2.61E-01 (3.74E-04) (失败) | 1+3 (0.02) | 4.40E-16 (2.42E-13) |
| 例 49 悬链线 [26] (n = 501, m = 166) | 1+397 (725.98) | 2.95E+01 (4.30E+05) (失败) | 85 (4.30) | 2.09E+08 (1.78E+00) (失败) | 7092+9790 (101.23) | 2.00E-06 (2.00E-11) |
| 考试 50 broydn3d [26] (n = 1000, m = 1000) | 2+0 (0.40) | 0 (3.72E-08) | 5 (1.59) | 5.08E-11 (8.88E-16) | 5+2 (0.31) | 1.70E-08 (1.56E-05) |
| 考试. 51 hs007 [26] (n = 2, m = 1) | 6+13 (0.057) | 2.39E-06 (0.00E+00) | 9 (0.0090) | 6.11E-08 (3.47E-12) | 20+21 (0.13) | 1.56E-05 (3.70E-07) |
| 考试 52 hs008 [26] (n = 2, m = 2) | 3+0 (0.0055) | 0.00E+00 (5.12E-08) | 5 (0.010) | 1.32E-06 (1.26E-10) | 4+1 (0.001) | 4.00E-11 (2.69E-07) |
| 考试 53 hs009 [26] (n = 2, m = 1) | 1+6 (0.014) | 3.84E-06 (0.00E+00) | 5 (0.0061) | 4.33E-06 (1.78E-15) | 5+6 (0.001) | 2.80E-12 (2.27E-13) |
| 考试. 54 gouldqp2 [26] (n = 699, m = 349) | 1+35 (3.038) | 8.01E-06 (2.00E-11) | 99 (7.36) | 2.19E-04 (4.62E-12) | 1400+1750 (0.88) | 1.90E-06 (1.14E-13) |
| 考试. 55 optcdeg2 [26] (n = 1202, m = 800) | 1+52 (17.73) | 8.79E-06 (7.71E-07) | 87 (25.77) | 8.62E-06 (3.13E-10) | 58+459 (0.297) | 1.30E-06 (1.14E-13) |
| Exam. 56 hs100lnp [26] (n = 7, m = 2) | 1+14 (0.18) | 4.26E-06 (1.42E-14) | 15 (0.32) | 7.32E-04 (8.11E-08) | 26+50 (0.218) | 1.10E-09 (2.67E-12) |
| 考试。57 hs111 [26] (n = 10, m = 3) | 4+29 (0.21) | 1.63E-02 (4.38E-05) (失败) | 8 (0.15) | 1.21E-01 (3.64E-03) (失败) | 50+57 (0.26) | 4.70E-07 (1.40E-08) |
| 考试 58 阅读1 [26] (n = 2002, m = 1000) | 1+47 (243.60) | 4.17E+09 (2.07E+13) (失败) | 99 (258.70) | 3.02E+02 (4.34E+03) (失败) | 827+38923 (205.094) | 1.50E+04 (3.22E+07) (失败) |
| Exam. 59 hs100lnp [26] (n = 4, m = 1) | 1+90 (1.40) | 9.91E-06 (1.32E-09) | 3 (1.28) | 4.60E-07 (8.53E-14) | 5+8 (0.297) | 4.30E-11 (1.14E-13) |
| 例6 读取2 [26] (n = 3003, m = 2002) | 1+394 (14094.66) | 2.53E-05 (5.45E-09) | 28 (3975.68) | 2.53E-05 (7.64E-03) | 2600+2650 (0.672) | 0.00E+00 (1.01E-10) |
| 考试 61 rk23 [26] (n = 17, m = 11) | 2+394 (0.38) | 1.00E+00 (2.57E-11) (失败) | 46 (0.1001) | 3.05E+04 (1.94E+01) (失败) | 0+3 (0.001) | 1.50E-05 (5.00E-01) (失败) |
| 考试 62 hs046 [26] (n = 5, m = 2) | 1+32 (0.038) | 8.84E-06 (6.23E-09) | 25 (0.032) | 3.01E-06 (1.42E-08) | 39+42 (0.234) | 1.90E-06 (2.11E-10) |
| 考试 63 hs099 [26] (n = 23, m = 18) | 1+32 (0.4808) | 8.84E-06 (6.23E-09) | 25 (0.110) | 3.01E-06 (1.42E-08) | 14+23 (0.016) | 1.30E-05 (4.51E-04) |
| 例64 hs0991 [26] (n = 31, m = 26) | 4+0 (0.039) | 8.81E+02 (1.40E-08) (失败) | 7 (0.086) | 1.06E+06 (9.90E+02) (失败) | 14+23 (0.015) | 9.00E-06 (3.15E-04) |
| 例6 65 hs99exp [26] (n = 31, m = 21) | 5+50 (0.92) | 2.91E+05 (7.05E+06) (失败) | 25 (0.067) | 4.03E+13 (8.85E+12) (失败) | 30+136 (0.016) | 3.80E-08 (3.37E-08) |

表 4：Rcm、fmincon 和 SNOPT 在具有$n = 2000, m = 10$ 的大规模问题上的数值结果。

| 问题 | Rcm | | fmincon | | SNOPT | |
|---|---|---|---|---|---|---|
| | $itc_G + itc_R$ (时间) | KKT $\|\|$约束$(x^{it})\|\|_\infty$ | 步数 (时间) | KKT $\|\|$约束$(x^{it})\|\|_\infty$ | 主要+ICTS (时间) | KKT $\|\|$约束$(x^{it})\|\|_\infty$ |
| 66. Trid函数 [26] (n = 2000, m = 10) | 1+12 (1.10) | 8.23e-06 (3.55e-15) | 13 (7.72) | 4.88e-06 (1.30e-11) | 4+1995 (3.48) | 7.20e-07 (0) |
| 67. 格里旺克函数 [26] (n = 2000, m = 10) | 6+10 (3.59) | 7.02e-06 (4.98e-15) | 99 (69.52) | 69.56 (1.12e-03) (失败) | 12+2013 (18.03) | 2.000e-07 (9.96e-12) |
| 68. Dixon-Price 函数 [26] (n = 2000, m = 10) | 5+9 (1.05) | 5.01e-06 (7.72e-12) | 25 (19.37) | 2.45e-06 (3.17e-08) | 31+2030 (44.95) | 9.20e-08 (2.38e-10) |
| 69. Rosenbrock 函数 [26] (n = 2000, m = 10) | 5+14 (1.50) | 5.21e-06 (1.10e-08) | 15 (7.73) | 1.44e-06 (2.96e-06) | 401+5952 (549.72) | 1.00e-06 (5.97e-11) |
| 70. 三角函数 [26] (n = 2000, m = 10) | 1+12 (3.68) | 8.58e-06 (1.43e-12) | 53 (45.72) | 5.98e-05 (1.32e-10) | 97+4126 (425.56) | 2.00e-06 (2.00e+03) (失败) |
| 71. 奇异布罗伊登函数 [26] (n = 2000, m = 10) | 10+34 (3.72) | 7.22e-06 (8.45e-07) | 32 (18.91) | 4.11e-06 (4.91e-11) | 211+2320 (321.55) | 1.500e-06 (2.11e+00) (失败) |
| 72. 扩展 Powell 奇异函数 [26] (n = 2000, m = 10) | 12+35 (9.30) | 9.28e-06 (4.44e-16) | 27 (14.53) | 7.14e-06 (1.43e-13) | 33+2025 (41.11) | 1.30e-06 (6.11e-08) |
| 73. 三对角系统函数 [26] (n = 2000, m = 10) | 1+28 (2.25) | 5.99e-06 (8.86e-15) | 64 (37.47) | 8.08e-06 (1.94e-10) | 5+1996 (8.89) | 4.60e-07 (0) |
| 74. 离散边值函数 [26] (n = 2000, m = 10) | 1+7 (5.10) | 4.05e-06 (4.24e-22) | 3 (4.54) | 4.01e-06 (5.59e-14) | 6+1997 (10.80) | 1.40e-07 (1.88e-09) |
| 75. 布罗伊登三对角函数 [26] (n = 2000, m = 10) | 2+9 (0.92) | 8.38e-07 (2.15e-10) | 13 (7.38) | 2.46e-06 (1.48e-08) | 68+2072 (93.66) | 1.40e-06 (1.35e+00) (失败) |
| 76. 扩展伍德函数 [26] (n = 2000, m = 10) | 11+14 (1.58) | 1.20e-06 (1.08e-06) | 58 (38.54) | 6.27e-08 (3.48e-09) | 23+2013 (35.87) | 1.10e-07 (9.54e-11) |
| 77. 扩展悬崖函数 [26] (n = 2000, m = 10) | 6+7 (1.11) | 1.41e-06 (6.66e-16) | 9 (4.79) | 8.18e-06 (1.35e-14) | 8+2003 (11.91) | 2.00e-08 (8.26e-06) |
| 78. 扩展希伯特函数 [26] (n = 2000, m = 10) | 1+56 (4.95) | 9.91e-07 (2.71e-25) | 99 (49.24) | 1.10e+03 (8.42e+04) (失败) | 7+1997 (12.02) | 3.30e-07 (5.02e-09) |
| 79. 扩展 Maratos 函数 [26] (n = 2000, m = 10) | 10+6 (1.06) | 1.76e-06 (1.42e-14) | 99 (63.59) | 3.31e+03 (3.40e+02) (失败) | 19+2014 (27.06) | 5.20e-08 (2.77e-05) |
| 80. 扩展 Psc1 函数 [26] (n = 2000, m = 10) | 7+8 (2.95) | 5.93e-06 (2.22e-16) | 19 (13.39) | 3.77e-05 (7.77e-16) | 11+2006 (18.78) | 8.20e-09 (5.58e-08) |
| 81. 扩展二次惩罚二次规划 1 函数 [26] (n = 2000, m = 10) | 1+9 (1.17) | 2.61e-06 (4.18e-41) | 6 (3.95) | 3.17e-07 (3.52e-10) | 8+1998 (13.36) | 2.10e-07 (0) |
| 82. 扩展二次罚 QP 2 函数 [26] (n = 2000, m = 10) | 1+6 (2.58) | 2.32e-06 (6.07e-39) | 5 (5.04) | 1.73e-04 (1.60e+10) | 8+1998 (17.22) | 2.80e-08 (0) |
| 83. 扩展 TET 函数 [26] (n = 2000, m = 10) | 7+6 (3.34) | 1.77e-06 (0) | 9 (6.67) | 5.15e-06 (2.22e-16) | 16+2011 (26.36) | 2.30e-09 (1.27e-08) |
| 84. EG2 函数 [26] (n = 2000, m = 10) | 1+6 (2.72) | 5.10e-06 (2.95e-07) | 6 (5.93) | 7.18e-06 (1.89e-07) | 6+1996 (12.55) | 1.70e-07 (1.14e-05) |
| 85. 扩展 BD1 函数 [26] (n = 2000, m = 10) | 5+6 (1.46) | 1.35e-06 (0) | 7 (4.29) | 8.82e-06 (2.22e-15) | 8+2003 (12.09) | 1.70e-06 (9.65e-08) |

表 5：Rcm、fmincon 和 SNOPT 在 n = 2000,{m } = 1000 的大规模问题上的数值结果。

| 问题 | Rcm | | fmincon | | SNOPT | |
|---|---|---|---|---|---|---|
| | $itc_G+itc_R$ (时间) | KKT $\|\|$约束（$x^{it}$）$\|\|_\infty$ | 步数 (时间) | KKT $\|\|$约束（$x^{it}$）$\|\|_\infty$ | 主要+ICTS (时间) | KKT $\|\|$约束（$x^{it}$）$\|\|_\infty$ |
| 86. Trid函数 [78] （n = 2000, m = 1000） | 1+1 (0.90) | 2.65e-06 (1.80e-06) | 7 (17.97) | 1.98e-05 (5.82e-11) | 0+1 (0.02) | 1.40e-13 (3.64e-12) |
| 87. 格里旺克函数 [78] (n = 2000, m = 1000) | 14+6 (8.66) | 4.07e-06 (3.30e-12) | 59 (1.07e+02) | 5.21e-06 (7.70e-164) | 32+2035 (84.20) | 1.40e-06 (4.64e-13) |
| 88. Dixon-Price 函数 [78] (n = 2000, m = 1000) | 8+11 (7.28) | 2.34e-06 (1.01e-07) | 13 (17.11) | 7.09e-07 (7.41e-07) | 219+1254 (41.47) | 8.20e-07 (1.82e-09) |
| 89. Rosenbrock 函数 [61] (n = 2000, m = 1000) | 6+14 (7.60) | 7.02e-06 (1.70e-08) | 16 (20.98) | 4.59e-06 (2.11e-05) | 276+2727 (55.38) | 7.00e-08 (2.19e-08) |
| 90. 三角函数 [61] (n = 2000, m = 1000) | 8+7 (7.67) | 9.74e-06 (5.58e-11) | 94 (1.23e+02) | 2.45e-02 (4.80e-11) （失败） | 245+7140 （1003.45） | 5.90e+02 （1.28e+03） （失败） |
| 91. 奇异布罗伊登函数 [39] (n = 2000, m = 1000) | 10+15 (9.97) | 8.38e-06 (1.92e-06) | 29 (39.45) | 3.25e-05 (3.24e-12) | 5413+84179 (91.94) | 1.20e-03 (2.16e+00) (失败) |
| 92. 扩展 Powell 奇异函数 [61] (n = 2000, m = 1000) | 12+10 (7.89) | 8.45e-06 (2.05e-14) | 16 (21.15) | 4.34e-06 (2.95e-09) | 18+1270 (6.06) | 1.80e-06 (1.62e-05) |
| 93. 三对角系统函数 [39] (n = 2000, m = 1000) | 1+22 (8.64) | 9.80e-06 (1.55e-15) | 27 (36.08) | 9.80e-06 (1.34e-08) | 4+1006 (1.89) | 1.20e-07 (0) |
| 94. 离散边值函数 [39] (n = 2000, m = 1000) | 1+7 (8.78) | 2.51e-06 (6.94e-18) | 6 (12.44) | 7.90e-07 (2.76e-14) | 5+1007 (2.28) | 5.20e-07 (2.96e-07) |
| 95. 布罗伊登三对角函数 [39] (n = 2000, m = 1000) | 2+10 (5.04) | 4.43e-06 (5.46e-09) | 14 (18.19) | 4.43e-06 (6.80e-09) | 11+15 (0.03) | 9.40e-14 (1.41) （失败） |
| 96. 扩展伍德函数 [1] (n = 2000, m = 1000) | 11+6 (6.58) | 1.29e-06 (0) | 58 (38.54) | 6.27e-08 (3.48e-09) | 13+1013 (4.77) | 1.40e-06 (8.53e-05) |
| 97. 扩展悬崖函数 [1] (n = 2000, m = 1000) | 6+7 (5.61) | 1.41e-06 (6.66e-16) | 13 (16.95) | 4.01e-09 (6.66e-16) | 12+1512 (3.34) | 3.30e-09 (5.82e-05) |
| 98. 扩展希伯特函数 [1] (n = 2000, m = 1000) | 1+43 (32.32) | 1.55e-06 (2.96e-23) | 99 (1.24e+02) | 2.03e+04 (3.09e+04) （失败） | 6+1006 (2.09) | 1.60e-07 （5.02e-09） |
| 99. 扩展 Maratos 函数 [1] （n = 2000, m = 1000） | 10+6 (6.31) | 2.00e-06 (1.42e-14) | 90 (2.01e+02) | 2.52e-07 (1.42e-14) | 37+1537 (8.81) | 2.80e-07 (5.34e-05) |
| 100. 扩展 Psc1 函数 [1] (n = 2000, m = 1000) | 7+8 (7.49) | 5.94e-06 (5.55e-16) | 48 (68.38) | 3.95e-05 (8.21e-12) | 12+1512 (2.34) | 6.40e-07 (2.43e-13) |
| 101. 扩展二次惩罚二次规划 1 函数 [1] (n = 2000, m = 1000) | 6+5 (4.72) | 6.86e-06 (8.31e-34) | 10 (14.46) | 1.85e-06 (8.40e-10) | 12+1012 (98.39) | 1.20e-08 (7.88e-06) |
| 102. 扩展二次罚 QP 2 函数 [1] (n = 2000, m = 1000) | 6+6 (6.19) | 2.20e-06 (6.62e-35) | 11 (36.49) | 1.22e-06 (1.99e-12) | 89+1094 (261.47) | 1.60e+01 (0.32) （失败） |
| 103. 扩展 TET 函数 [1] (n = 2000, m = 1000) | 7+6 (7.19) | 4.67e-06 (0) | 11 (16.70) | 6.61e-07 (0) | 18+1518 (4.88) | 5.90e-07 (7.02e-14) |
| 104. EG2 函数 [1] (n = 2000, m = 1000) | 1+6 (5.06) | 5.11e-06 (2.95e-07) | 6 (12.34) | 7.72e-06 (1.59e-07) | 6+1006 (2.81) | 3.60e-08 (1.89e-08) |
| 105. 扩展 BD1 函数 [1] (n = 2000, m = 1000) | 5+6 (5.08) | 1.89e-06 (0) | 9 (13.34) | 2.50e-07 (0) | 9+1509 (2.58) | 2.30e-07 (6.53e-05) |

表 6：Rcm、fmincon 和 SNOPT 在大规模问题上的数值结果，$n = 2000$，$m = 1999$。

| 问题 | Rcm | | fmincon | | SNOPT | |
|---|---|---|---|---|---|---|
| | $itc_G + itc_R$ (时间) | KKT $\|\|$约束$(x^{it})\|\|_\infty$ | 步数 (时间) | KKT $\|\|$约束$(x^{it})\|\|_\infty$ | 主要+ICTS (时间) | KKT $\|\|$约束$(x^{it})\|\|_\infty$ |
| 106. Trid函数 [78] (n = 2000, m = 1999) | 1+1 (1.95) | 9.21e-06 (1.67e-08) | 2 (12.19) | 2.54e-06 (2.33e-10) | 0+1 (0.02) | 3.70e-12 (7.28e-12) |
| 107. 格里旺克函数 [78] (n = 2000, m = 1999) | 21+8 (12.67) | 9.48e-06 (7.76e-17) | 32 (1.86e+02) | 5.47e-04 (2.53e-15) | 30+2037 (121.20) | 1.70e-07 (4.11e-10) |
| 108. Dixon-Price 函数 [78] (n = 2000, m = 1999) | 8+6 (5.32) | 2.20e-06 (1.42e-09) | 97 (5.44e+02) | 1.25e+03 (1.95e+02) (失败) | 11+16 (0.03) | 1.10e-08 (7.46e-07) |
| 109. Rosenbrock 函数 [61] (n = 2000, m = 1999) | 6+27 (21.10) | 2.25e-06 (6.20e-09) | 20 (64.61) | 1.39e-06 (1.07e-06) | 253+223 (3.19) | 6.30e-05 (7.58e-07) |
| 110. 三角函数 [61] (n = 2000, m = 1999) | 1+13 (15.29) | 1.24e-06 (1.13e-15) | 15 (68.47) | 1.41e-03 (2.89e-13) (失败) | 96+10417 (1003.13) | 1.60e+03 (9.71e+02) (失败) |
| 111. 奇异布罗伊登函数 [39] (n = 2000, m = 1999) | 10+6 (19.26) | 1.84e-06 (6.40e-07) | 74 (2.87e+02) | 28.52 (9.41e-13) (失败) | 49586+606083 (480.78) | 3.70e-03 (2.16e+00) (失败) |
| 112. 扩展 Powell 奇异函数 [61] (n = 2000, m = 1999) | 12+26 (21.14) | 5.45e-06 (6.38e-11) | 27 (1.04e+02) | 6.46e-06 (7.70e-12) | 23+525 (0.22) | 4.00e-07 (5.30e-08) |
| 113. 三对角系统函数 [39] (n = 2000, m = 1999) | 1+15 (9.00) | 9.28e-06 (1.55e-15) | 35 (1.08e+02) | 1.28e-02 (4.88e-15) | 1+3 (0.02) | 6.10e-18 (0) |
| 114. 离散边值函数 [39] (n = 2000, m = 1999) | 1+7 (10.60) | 7.57e-06 (8.32e-17) | 17 (70.40) | 1.16e-06 (2.49e-16) | 7+10 (0.03) | 1.60e-07 (8.72e-07) |
| 115. 布罗伊登三对角函数 [39] (n = 2000, m = 1999) | 2+7 (6.32) | 2.47e-06 (7.15e-10) | 11 (33.86) | 2.77e-06 (2.46e-06) | 11+16 (0.11) | 1.20e-08 (1.41e+00) (失败) |
| 116. 扩展伍德函数 [1] (n = 2000, m = 1999) | 11+5 (9.35) | 3.69e-06 (3.79e-09) | 37 (1.46e+02) | 2.56e-08 (4.14e-11) | 54+55 (0.23) | 7.00e-09 (3.30e-08) |
| 117. 扩展悬崖函数 [1] (n = 2000, m = 1999) | 6+7 (8.60) | 1.41e-06 (6.66e-16) | 8 (45.13) | 1.29e-08 (1.13e-13) | 9+1009 (0.09) | 1.30e-08 (7.52e-07) |
| 118. 扩展希伯特函数 [1] (n = 2000, m = 1999) | 1+10 (7.61) | 8.09e-06 (3.28e-42) | 5 (16.01) | 1.90e-06 (2.02e-13) | 6+12 (0.03) | 1.40e-06 (5.02e-09) |
| 119. 扩展 Maratos 函数 [1] (n = 2000, m = 1999) | 10+14 (12.98) | 7.99e-06 (2.20e-06) | 99 (3.08e+02) | 73.00 (49.68) (失败) | 27+1027 (0.23) | 1.50e-06 (6.32e-10) |
| 120. 扩展 Psc1 函数 [1] (n = 2000, m = 1999) | 7+14 (13.73) | 2.20e-06 (3.33e-16) | 14 (43.92) | 9.75e-06 (4.49e-11) | 12+1012 (0.13) | 7.50e-07 (7.61e-07) |
| 121. 扩展二次惩罚二次规划 1 函数 [1] (n = 2000, m = 1999) | 13+19 (17.29) | 9.63e-06 (7.22e-07) | 53 (1.64e+02) | 8.95e-02 (1.74e-61) (失败) | 16+17 (121.86) | 1.30e-09 (6.16e-09) |
| 122. 扩展二次罚 QP 2 函数 [1] (n = 2000, m = 1999) | 14+16 (11.52) | 2.42e-06 (1.27e-08) | 12 (44.11) | 4.27e-06 (8.08e-10) | 1+2 (605.25) | 3.00e-04 (1.59e+10) (失败) |
| 123. 扩展 TET 函数 [1] (n = 2000, m = 1999) | 7+7 (10.34) | 5.45e-06 (5.50e-12) | 12 (38.43) | 4.53e-07 (1.26e-07) | 21+1021 (0.17) | 1.30e-06 (9.35e-08) |
| 124. EG2 函数 [1] (n = 2000, m = 1999) | 3+6 (7.80) | 1.30e-06 (3.68e-13) | 8 (26.53) | 1.80e-06 (3.68e-13) | 10+11 (0.27) | 2.50e-08 (2.99e-10) |
| 125. 扩展 BD1 函数 [1] (n = 2000, m = 1999) | 5+21 (14.41) | 3.88e-06 (2.03e-09) | 13 (42.83) | 8.68e-08 (8.60e-10) | 14+1014 (0.19) | 6.50e-07 (6.60e-06) |

表 7：Rcm、fmincon 和 SNOPT 求解失败的问题数。

| | Rcm | fmincon | SNOPT |
|---|---|---|---|
| 失败问题数 | 8 | 28 | 21 |
| 失败概率 | $\frac{8}{125}$ (6.4%) | $\frac{28}{125}$ (22.4%) | $\frac{21}{125}$ (16.8%) |
| 快速问题数 | 48 | 3 | 72 |
| 快速问题的概率 | $\frac{48}{125}$ (38.4%) | $\frac{3}{125}$ (2.4%) | $\frac{72}{125}$ (57.6%) |

## 5 结论

本文提出了一种结合信赖域更新策略（Rcm）的非线性等式约束优化问题正则化延拓法。具体而言，在适定阶段，我们使用正则化拟牛顿矩阵的逆作为预条件子，以提高计算效率；在病态阶段，我们使用海森矩阵的正则化双侧投影的逆作为预条件子，以增强鲁棒性。由于 Rcm 在每次迭代中只需求解一个线性方程组，而序列二次规划（SQP）在每次迭代中都需要求解一个二次规划子问题，因此 Rcm 比 SQP 更快。数值结果也表明，Rcm 比 SQP 更鲁棒、速度更快（SQP 包括 MATLAB2020a 环境中内置的子程序 fmincon.m [60,72] 以及 GAMS v28.2（2019年）环境中执行的子程序 SNOPT [29,30] [23]）。对于大规模问题，Rcm 的计算时间约为 fmincon.m 的三分之一。因此，Rcm 是一种用于等式约束优化问题的交替求解器；正则化延拓法在不等式约束优化问题和正交矩阵约束问题中的应用也值得进一步探索 [84]。

## 致谢

本研究部分得到了国家自然科学基金委员会项目（资助号：61876199）、华为技术有限公司项目（资助号：YBWL2011085）以及华为技术有限公司创新研究计划项目（资助号：YJCB2011003HI）的资助。作者感谢两位匿名审稿人提出的意见和建议，这些意见和建议极大地改善了本文的表述。

**利益冲突/竞争性利益：** 不适用。

**数据和材料的可用性（数据透明度）：** 如有需要，我们将提供测试数据。

**代码可用性（软件应用程序或定制代码）：** 如有需要，我们将提供代码。

## 参考文献

1. N. Andrei，无约束优化测试函数集，Adv Model Optim**10** (2008), 147-161. 2. E. P. Adorio 和 U. P. Diliman，*MVF*——用于无约束全局优化的 *C* 语言多变量测试函数库，http://www.geocities.ws/eadorio/mvf.pdf，2005. 3. E. L. Allgower 和 K. Georg，数值延拓方法导论，SIAM，费城，宾夕法尼亚州，2003. 4. U. M. Ascher 和 L. R. Petzold，常微分方程与微分代数方程的计算机方法，SIAM，费城，宾夕法尼亚州，1998. 5. C. G. Broyden，一类双秩极小化算法的收敛性，J Inst Math Appl **6** (1970), 76-90.

6. A. A. Brown 和 M. C. Bartholomew-Biggs，约束优化的常微分方程方法与序列二次规划方法，《优化理论与应用杂志》62 (1989), 371-386. 7. K. E. Brenan、S. L. Campbell 和 L. R. Petzold，《微分代数方程初值问题的数值解》，SIAM，费城，宾夕法尼亚州，1996. 8. J. M. Bioucas-Dias 和 M. A. T. Figueiredo，用于约束稀疏回归的交替方向算法：高光谱解混应用，2010年第2届高光谱图像与信号处理：遥感发展研讨会，

2010, 1 4 http://doi.org/10.1109/ - ， 。 WHISPERS.2010.5594963

9. J. C. Butcher 、 Z. Jackiewicz，常微分方程高阶对角隐式多阶段积分方法的构造，Appl Numer Math 27 (1998), 1-12. 10. R. Byrd、J. Nocedal 和 Y. X. Yuan，一类拟牛顿法在凸问题上的全局收敛性，SIAM 数值分析期刊 24 (1987), 1171-1189. 11. F. Caballero、L. Merino、J. Ferruz 和 A. Ollero，用于中高空飞行无人机的基于视觉的里程计与同步定位与建图，J Intell Robot Syst 54 (2009), 137-161. 12. T. S. Coffey、C. T. Kelley 和 D. E. Keyes，伪瞬态延拓与微分代数方程，SIAM J Sci Comput 25 (2003), 553-569. 13. A. R. Conn、N. Gould 和 Ph. L. Toint，《信赖域方法》，SIAM，费城，美国，2000. 14. M. T. Chu 和 M. M. Lin，中心路径及其变体的动力系统刻画——再探，SIAM J Appl Dyn Syst 10 (2011), 887-905. 15. A. d' Aspremont、L. El Ghaoui、M. Jordan、G. R. Lanckriet，利用半定规划求解稀疏主成分分析的直接形式，《SIAM评论》49 (2007), 434-448. 16. E. D. Dolan 和 J. J. Mor´e，利用性能剖面对优化软件进行基准测试，《数学规划》91 (2002), 201-213. 17. A. V. Fiacco 和 G. P. McCormick，《非线性规划：序列无约束最小化技术》，SIAM，1990. 18. R. Fletcher，一种变度量算法的新方法，Comput J 13 (1970), 317-322. 19. R. Fletcher 和 M. J. D. Powell，一种快速收敛的极小化下降法，Comput J 6 (1963), 163-168. 20. J. M. Fadili、J. L. Starck，用于稀疏恢复优化问题的单调算子分裂，IEEE国际图像处理会议，11月

2009， 1461 1464 http://doi.org/10.1109/ICIP.2009. 开 5414555

罗，埃及，- ，。 21. M. A. T. Figueiredo 和 J. M. Bioucas-Dias，使用交替方向优化复原泊松图像，《IEEE图像处理汇刊》19 (2010), 3133-3145. 22. P. A. Forero、A. Cano 和 G. B. Giannakis，基于共识的分布式支持向量机，《机器学习研究杂志》11 (2010), 1663-1707. 23. GAMS v28.2，GAMS公司， https://www.gams.com/，2019年. 24. B. S. Goh，具有等式约束的优化问题的近似最速下降法，《优化理论与应用杂志》148 (2011), 505-527. 25. D. Goldfarb，通过变分方法导出的一族变度量更新，《数学规划》24 (1970), 23-26. 26. N. I. M. Gould、D. Orban 和 P. L. Toint，CUTEst：用于数学优化的安全线程约束与无约束测试环境，《计算优化与应用》60 (2015), 545-557. 27. G. H. Golub 和 C. F. Van Loan，《矩阵计算》第4版，约翰斯·霍普金斯大学出版社，巴尔的摩，Mayryland，2013. 28. P. E. Gill、W. Murray 和 M. H. Wright，《实用优化》，学术出版社，伦敦，1981. 29. P. E. Gill、W. Murray 和 M. A Saunders，SNOPT：用于大规模约束优化的序列二次规划算法，《SIAM评论》47, (2005), 99-131. 30. P. E. Gill、W. Murray 和 M. A Saunders，SQOPT第7版用户指南：用于大规模线性规划和二次规划的软件，2006. 31. S. P. Han，一种非线性规划全局收敛法，《优化理论与应用杂志》22 (1977), 297-307. 32. P. C. Hansen，《正则化工具：用于离散不适定问题分析与求解的 MATLAB 工具包》，Numer Algorithms 6 (1994), 1-35. 33. U. Helmke 和 J. B. Moore，《优化与动力系统》第2版，施普林格出版社，伦敦，1996. 34. D. J. Higham，信赖域算法与时间步长选择，SIAM 数值分析期刊 37 (1999), 194-210. 35. W. Hock 和 K. Schittkowski，27种非线性规划程序的比较性能评估，Computing 30 (1983), 335-358，网址： https://doi.org/10.1007/BF02242139。

*36. Z. Jackiewicz* 和 *S. Tracogna*，常微分方程的一类两步 *Runge-Kutta* 方法，《*SIAM* 数值分析期刊》*32 (1995)*，*1390-1427*。*37. C. T. Kelley*、*L.-Z. Liao*、*L. Qi*、*M. T. Chu*、*J. P. Reese* 和 *C. Winton*，投影伪瞬态延拓，《*SIAM* 数值分析期刊》*46 (2008)*，*3071-3083*。*38. J. H. Lee*、*Y. M Jung*、*Y. X. Yuan* 和 *S. Yun*，一种用于等式约束优化的子空间 *SQP* 方法，《计算优化与应用》*74 (2019)*，*177-194*。*39. L. Lukˇsan*，用于大型稀疏非线性方程组的非精确信赖域方法，《优化理论与应用杂志》*81*，*569-590 (1994)*。*40. K. Levenberg*，一种求解某些最小二乘问题的方法，《应用数学季刊》*2 (1944)*，*164-168*。*41. L.-Z. Liao*、*H. D. Qi* 和 *L. Q. Qi*，神经动力学优化，《全局优化杂志》*28 (2004)*，*175-195*。*42. X.-W. Liu* 和 *Y.-Y. Yuan*，一种无需罚函数或过滤器的非线性等式约束优化序列二次规划方法，《*SIAM* 优化期刊》*21 (2011)*，*545-571*。*43. Z. S. Lu*、*K. K. Pong* 和 *Y. Zhang*，一种求解 *Dantzig* 选择器的交替方向法，《计算统计与数据分析》

**56** 2012 4037 4046 https://doi.org/10.1016/j.csda.2012.04. ( )，-，。*44.*

019
*X.-L. Luo*，结合线搜索技术的单对角隐式 *Runge-Kutta* 无约束优化方法，《计算数学杂志》*23 (2005)*，*153-164*。*45. X.-L. Luo*、*L.-Z. Liao* 和 *H.-W. Tam*，*Levenberg-Marquardt* 方法的收敛性分析，《优化方法与软件》*22 (2007)*，*659-678*。*46. S.-T. Liu* 和 *X.-L. Luo*，一种基于 *Rayleigh* 商梯度流的极端及内部特征值问题求解方法，《线性代数及其应用》*432 (2010)*，*1851-1863*。*47. X.-L. Luo*，一种用于最小特征值问题的微分代数方程动力学方法，《计算机科学杂志》*3 (2012)*，*113-119*。*48. X.-L. Luo*、*J.-R. Lin* 和 *W.-L. Wu*，一种用于大规模广义特征值问题的预测校正动态方法，《抽象与应用分析》(

2013 845459, 1 8 http://dx.doi.org/ )，文章编号 -，。

10.1155/2013/845459
*49. X.-L. Luo*、*J.-H. Lv* 和 *G. Sun*，采用可信时间步进方案的延拓方法，用于含噪数据的线性约束优化，《优化工程》 **23** 2022 329 360 http://doi.org/ ( )，-，。

10.1007/s11081-020-09590-z
*50. X.-L. Luo*、*H. Xiao* 和 *J.-H. Lv*，采用残差信赖域时间步进方案的延拓牛顿方法，用于非线性方程，《数值算法》 **89** 2022 223 247 http://doi.org/ ( )，-，。

10.1007/s11075-021-01112-x
*51. X.-L. Luo* 和 *Y.-Y. Yao*，用于含噪数据线性规划的原始-对偶路径跟踪方法及信赖域更新策略，《计算数学杂志》 **40** 2022 760 780 http://doi.org/10. ( )，-，。

4208/jcm.2101-m2020-0173
*52. X.-L. Luo*、*H. Xiao*、*J.-H. Lv* 和 *S. Zhang*，用于无约束优化的显式伪瞬态延拓及信赖域更新策略，《应用数值数学》*165 (2021)*，*290-302*，

http://doi.org/10.1016/j.apnum.2021.02.019 。*53. X.-L. Luo* 和 *H. Xiao*，欠定系统的广义延拓牛顿方法及信赖域更新策略，《科学计算杂志》

**88** 2021 56， 1 22 http://doi. ( )，文章页码 -，。

org/10.1007/s10915-021-01566-0
*54. X.-L. Luo* 和 *H. Xiao*，一种用于线性约束优化问题、带自适应时间步长控制的正则化延拓法，《应用数值数学》*181 (2022)*，*255-276*，在线发表于 https://doi.org/10.1016/j.apnum.2022.06.008。*55. X.-L. Luo*、*H. Xiao* 和 *S. Zhang*，采用消去技术的延拓牛顿方法，用于全局优化问题，*arXiv* 预印本见

http://arxiv.org/abs/2107.13864，或 *Research Square* 预印本见

https://doi.org/10.21203/rs.3.rs-1102775/v1，

30，2021 https://teacher.bupt.edu.cn/luoxinlong/zh_CN/zzcg/ 年7月。软件见

41406/list/index.htm
。*56. X.-L. Luo* 、 *S. Zhang* 和 *H. Xiao*，线性互补问题的残差正则化路径跟踪方法，*arXiv* 预印本见

http://arxiv.org/abs/2205.10727，第*1-30*页，*2022*年*5*月*21*日。*57. M. Ng*、*P. Weiss* 和 *X.-M. Yuan*，通过交替方向法求解约束总变差图像复原与重建问题，《*SIAM* 科学计算杂志》*32 (2010)*，*2710-2736*，

http://doi.org/10.1137/090774823 。*58. M. F. Mascarenhas*，精确线搜索下的 *BFGS* 方法在非凸目标函数上会失效，《数学规划》*99 (2004)*，*49-61*。

59. M.-W. Mak，《约束优化与支持向量机讲义》，http://www.eie.polyu.edu.hk/~mwmak/EIE6207/ContOpt-SVM-beamer.pdf 2019，。

60. MATLAB v9.8.0 (R2020a)，The MathWorks Inc.，http://www.mathworks.com，2020年。61. J. J. Mor´e、B. S. Garbow 和 K. E. Hillstrom，测试无约束优化软件，《ACM 软件交易》7 (1981)，17-41。62. D. Marquardt，一种非线性参数最小二乘估计算法，《SIAM 应用数学杂志》11 (1963)，431-441。63. N. Maculan 和 C. Lavor，一种用于测试分子势能全局极小化方法的函数，《数值算法》35 (2004)，287-300。64. E. H. Moore，论一般代数矩阵的倒数，《美国数学会公报（新系列）》26 (1920)，394-395。65. J. Nocedal 和 S. J. Wright，《数值优化》，施普林格出版社，柏林，1999。66. M. J. Osborne，《经济理论的数学方法》，https://mjo.osborne.economics.utoronto.ca/index.php/tutorial/index/1/mem 2016，。

67. P.-Q. Pan，等式约束优化的新常微分方程方法（2）：算法，《计算数学杂志》10 (1992)，129-146。68. R. A. Penrose，矩阵的广义逆，《剑桥哲学会数学学报》51 (1955)，406-413。69. M. J. D. Powell，一类极小化算法的收敛性质，载于：O.L. Mangasarian、R. R. Meyer 和 S. M. Robinson 编，《非线性规划 2》，学术出版社，纽约，第1-27页，1975。70. M. J. D. Powell，一种用于非线性约束优化计算的快速算法，载于：G. A. Watson 编，《数值分析》，施普林格出版社，柏林，第144-157页，1978。71. M. J. D. Powell，变度量方法在非线性约束优化计算中的收敛性，载于：O.L. Mangasarian、R. R. Meyer 和 S. M. Robinson 编，《非线性规划 3》，学术出版社，纽约，第27-63页，1978。72. K. Schittkowski，NLPQL：用于求解约束非线性规划问题的 Fortran 子程序，《运筹学年鉴》5 (1986)，485-500，见 https://doi.org/10.1007/BF02739235。73. J. Schropp，一种约束极小化的动力系统方法，《数值泛函分析与优化》21 (2000)，537-551。74. J. Schropp，指标二微分代数系统的单步与多步离散化及其在优化中的应用，《计算与应用数学杂志》150 (2003)，375-396。75. L. F. Shampine、I. Gladwell 和 S. Thompson，《用 MATLAB 求解常微分方程》，剑桥大学出版社，剑桥，2003。76. D. F. Shanno，用于函数极小化的拟牛顿法的条件数，《数学计算》24 (1970)，647-656。77. G. Steidl 和 T. Teuber，通过 Douglas-Rachford 分裂方法去除乘性噪声，《数学成像与视觉杂志》36 (2010)，168-184。78. S. Surjanovic 和 D. Bingham，仿真实验虚拟库：测试函数与数据集，取自 http://www.sfu.ca/~ssurjano，2020年1月。79. W. Y. Sun 和 Y. X. Yuan，《优化理论与方法：非线性规划》，施普林格出版社，纽约，2006。80. K. Tanabe，非线性规划中的一种几何方法，《优化理论与应用杂志》30 (1980)，181-210。81. A. N. Tikhonov，反问题的稳定性，《苏联科学院院刊》39 (1943)，176-179。82. A. N. Tikhonov 和 V. Y. Arsenin，《不适定问题的解》，约翰·威立父子出版公司，纽约—多伦多—伦敦，1977。83. R. Vanderbei、K. Lin、H. Liu 和 L. Wang，重新审视压缩感知：发挥单纯形方法与稀疏化方法的效率，《数学规划计算》

**8** 2016 253 269 http://doi.org/10.( )，-，。1007/s12532-016-0105-y

84. Z.-W. Wen 和 W.-T. Yin，一种带正交约束优化的可行方法，《数学规划》142 (2013)，397-434。85. R. B. Wilson，《凸规划单纯形方法》，博士论文，哈佛大学，1963。86. D. M. Witten、R. Tibshirani 和 T. Hastie，一种惩罚矩阵分解及其在稀疏主成分和典型相关分析中的应用，《生物统计学》10 (2009)，515-534。87. H. Yamashita，非线性规划的微分方程方法，《数学规划》18 (1980)，155-168。88. Y. Yuan，信赖域算法的最新进展，《数学规划》151 (2015)，249-281。