Алгоритмическая Выпуклая Оптимизация

Ю.Е. Нестеров

26 декабря 2013г.

Диссертация на соискание ученой степени
доктора физико-математических наук
по специальности 01.01.07 - вычислительная математика

Мотивировка (2002-2007)

- К началу 2000х стало ясно что границы Теории Сложности можно обойти.
- Пример: полиномиальные алгоритмы внутренней точки.
- Способ: заглянуть в Черный Ящик.
- Вопрос: Как?

Цель: Разработка новых методов, превосходящих существующие в одном из следующих аспектов:

- вычисление дополнительных характеристик решения (прямо-двойственные ГМ);
- гарантиями глобальной эффективности (новые методы 2го порядка);
- сходятся быстрее, чем разрешает стандартная теория сложности (техника сглаживания).

Подобная эффективность ранее была недостижимой.

Средства: использование модели целевой функции

1. Геометрические модели. Обосновывается факт включения некоторого простого выпуклого множества в более сложный объект. Примеры:

- *Дикинский эллипсоид.*
- *Аппроксимации 1го и 2го порядка.*

2. Алгебраические модели. В них описывается конкретный способ представления целевой функции. Пример:

$$f(x) \quad = \quad \max_{u \in Q_2} \{ \langle Ax, u \rangle - \phi(u) \}.$$

3. Структурные модели. В них описывается структура целевой функции. Обычно выделяются два взаимодействующих объекта: $f(x) + \Psi(x)$, $F(A(x))$, и т.д.

Содержание, I

Глава 1: Предварительные результаты

1.1 Классификация выпуклых функций.

1.2. Методы гладкой минимизации первого порядка.

1.3. Самосогласованные функции и барьеры.

Глава 2: Современные субградиентные методы

2.1. Прямо-двойственные методы для негладких задач.

2.2. Барьерный субградиентный метод.

2.3. Градиентные методы минимизации составных функций.

Глава 3: Вариационные неравенства

3.1. Вариационные неравенства с гладким оператором.

3.2. Сильно монотонные операторы и квазивариационные неравенства.

Содержание, II

Глава 4: Методы второго порядка
4.1 Кубическая регуляризация метода Ньютона.
4.2. Ускоренная кубическая регуляризация.
4.3. Модифицированный метод Гаусса-Ньютона.

Глава 5: Техника сглаживания
5.1. Сглаживание для явной модели целевой функции.
5.2. Условие обратного зазора в негладкой выпуклой минимизации.
5.3. Техника сглаживания в полуопределенной оптимизации.

Глава 6: Оптимизация с относительной точностью
6.1. Однородная модель целевой функции.
6.2. Эллипсоидальная аппроксимация выпуклых тел.

2.1. Прямо-двойственные методы для негладких задач

Задача: $\min_{x \in Q} f(x)$, где

- Q - простое выпуклое замкнутое множество (возможно, неограниченное).
- f -выпуклая функция с ограниченными субградиентами.

Стандартный метод: $x_{k+1} = \pi_Q(x_k - h_k \nabla f(x_k))$, $h_k \approx \frac{R}{L\sqrt{k}}$.

Новый метод:

$$x_{k+1} = \arg\min_{x \in Q} \left[\frac{1}{k+1} \sum_{i=0}^{k} \langle \nabla f(x_i), x \rangle + \beta_k \|x - x_0\|^2 \right], \ \beta_k \approx \frac{L}{R\sqrt{k}}.$$

Преимущества: а) две управляющие последовательности. б) Возможность восстановления двойственного решения по решениям вспомогательной задачи.

Нужна структура $f(x)$.

2.2. Барьерный субградиентный метод

Задача: $\max\limits_{x \in Q} f(x)$, где

- Q - выпуклое множество, заданное самосогл. бар. $F(x)$.
- f -вогнутая функция. Введем $\|s\|_x^* \overset{\text{def}}{=} \langle s, [\nabla^2 F(x)]^{-1} s \rangle^{1/2}$.

Вспомогательная задача: пусть $x_0 = \arg\min\limits_{x \in Q} F(x)$. Найдем $u_\beta^\star(s) = \arg\max\limits_{x \in Q} \{\langle s, x - x_0 \rangle - \beta[F(x) - F(x_0)]\}$, $s \in E^*$, $\beta > 0$.

Метод: $s_0 = 0$. Выбираем β_k, λ_k и вычисляем

$$x_k = u_\beta^\star(s_k), \qquad s_{k+1} = s_k + \lambda_k \nabla f(x_k).$$

Параметры: если $\|\nabla f(x)\|_x^* \le M$ для любого $x \in Q$, то выберем $\lambda_k = 1$, $k \ge 0$, $\beta_0 = \beta_1$, $\beta_k = M \cdot \left(1 + \sqrt{\frac{k}{\nu}}\right)$, $k \ge 1$.

Скорость: $f^* - f(x_k) \le O(M\sqrt{\frac{\nu}{k+1}})$.

Пример: Если $\psi > 0$ на Q, то у $f = \ln \psi$ будет $M = 1$.

2.3. Градиентные методы для составных функций

Задача: $\min\limits_{x \in Q} [\, f(x) + \Psi(x) \,]$, где

- Q - выпуклое замкнутое множество,
- f -выпуклая гладкая функция,
- Ψ - выпуклая замкнутая простая функция.

Вспомогательная задача: Найти

$$\min_{x \in Q} \left\{ \langle s, x \rangle + \beta \|x - \bar{x}\|^2 + \Psi(x) \right\}$$

Результаты:

- Обычные градиентные методы сходятся как $O(1/k)$.
- Быстрые градиентные методы сходятся как $O(1/k^2)$.

Пример: $f(x) = \frac{1}{2}\|Ax - b\|^2 + \|x\|_1$.

3.1. Вариационные неравенства с гладким оператором

Задача: Найти $x^* \in Q \;:\quad \langle g(x), x - x^* \rangle \geq 0 \quad \forall x \in Q$ где

- Q - выпуклое замкнутое множество,
- g - монотонный липшицев оператор.

Мера близости: $f_D(x) = \max\limits_{y \in Q} \{ \langle g(y), x - y \rangle : \; \| \bar{x} - y \| \leq D \}$.

Обозначим $T_\beta(z, s) = \arg\max\limits_{x} \{ \langle s, x - z \rangle - \frac{\beta}{2} \| x - z \|^2 : \; x \in Q \}$.

Итерация: $s \to s_+$ где

$$\mathcal{E}_{\beta,\lambda}(s) = (x, y, s_+) = \quad \Leftrightarrow \quad \left\{ \begin{array}{lcl} x & = & T_\beta(\bar{x}, s), \\ y & = & T_\beta(x, -\lambda\, g(x)), \\ s_+ & = & s - \lambda\, g(y). \end{array} \right.$$

Теорема: Пусть $\beta = L$, $\lambda = 1$, $\tilde{y}_k = \frac{1}{k+1} \sum\limits_{i=0}^{k} y_i$. Тогда

$$f_D(\tilde{y}_k) \leq \frac{LD}{\sigma \cdot (k+1)}, \quad k \geq 0.$$

Пусть $\| g(x_1) - g(x_2) \|_* \leq M$ и $\beta_k = M \sqrt{\frac{k+1}{\sigma D}}$, тогда

$$f_D(\tilde{y}_k) \leq 2M \sqrt{\frac{D}{\sigma \cdot (k+1)}}.$$

3.2. Сильно монотонные операторы и квазивариационные неравенства

Сильная монотонность: $\langle g(x) - g(y), x - y \rangle \geq \mu \|x - y\|^2$

Мера близости: $f(x) = \sup\limits_{y \in Q} \left\{ \langle g(y), x - y \rangle + \frac{1}{2}\mu \|y - x\|^2 \right\}$.

Для $\beta > 0$ обозначим $\psi_y^\beta(x) = \langle g(y), y - x \rangle - \frac{1}{2}\beta \|x - y\|^2$,

$\Psi_k(x) = \sum\limits_{i=0}^{k} \lambda_i \psi_{y_i}^\mu(x)$, $S_k = \sum\limits_{i=0}^{k} \lambda_i$.

Обозначим через $\gamma = \frac{L}{\mu} \geq 1$ *число обусловленности* g.

Итерация (*): $(y_0 = \bar{x})$

$x_k = \arg\max\limits_{x \in Q} \Psi_k(x)$, $y_{k+1} = \arg\max\limits_{x \in Q} \psi_{x_k}^L(x)$, $\lambda_{k+1} = \frac{1}{\gamma} \cdot S_k$.

Теорема: Для $\tilde{y}_k = \frac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i y_i$ имеем

$$f(\tilde{y}_k) \leq f(\bar{x}) \cdot \gamma^2 \cdot \exp\left\{ -\frac{k}{\gamma+1} \right\}.$$

Квазиварaиционное неравенство:

Найти $x_* \in \mathcal{Q}(x_*)$: $\quad \langle g(x_*), y - x_* \rangle \geq 0, \quad \forall y \in \mathcal{Q}(x_*)$,
где т-м отображение $\mathcal{Q}$ принимает вып. и замк. значения.

Условие: $\|\pi_{\mathcal{Q}(x)}(z) - \pi_{\mathcal{Q}(y)}(z)\| \leq \alpha \|x - y\|, \quad \forall x, y, z \in E$.

Обозначим $\tilde{y}_N(Q, \hat{x})$ рез-т N-шагового Метода (*),
применяемого к g на Q начиная с $\bar{x} = \arg\max\limits_{x \in Q} \psi_{\hat{x}}^L(x)$.

Пусть $\alpha\gamma < 1$. Зададим $\hat{N} = \lfloor 2(\gamma + 1) \ln \frac{12\gamma}{1-\alpha\gamma} \rfloor + 1$.
Рассмотрим схему

Выбираем $u_0 \in E$. Для $k \geq 0$ повторяем:
$\hat{x}_k = \pi_{\mathcal{Q}(u_k)}(u_k), \quad u_{k+1} = \tilde{y}_{\hat{N}}(\mathcal{Q}(u_k), \hat{x}_k)$.

Теорема: Существует единственное решение x_* и
$\|u_k - x_*\| \leq \frac{1}{\delta} \cdot \exp\left\{-\frac{\delta}{2} k\right\} \cdot \|u_0 - T(u_0)\|$.

4.1. Кубическая регуляризация метода Ньютона

Метод: $x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$. Недостатки:

- Не различает максимум, минимум, седловую точку.
- Может расходится.
- Не всегда корректно определен.
- Нет глобальных оценок.

Модель: для $\|\nabla^2 f(x) - \nabla^2 f(y)\| \le L\|x-y\|$ определим

$$\xi_{2,x}(h) = f(x) + \langle \nabla f(x), h\rangle + \tfrac{1}{2}\langle \nabla^2 f(x) h, h\rangle + \tfrac{L}{6}\|h\|^3$$

Тогда $f(x+h) \le \xi_{2,x}(h)$, $\forall h \in R^n$, и можно применить

$$x_{k+1} \in x_k + \operatorname{Arg}\min_h \ \xi_{2,x_k}(h) \quad (\equiv T_L(x_k))$$

Преимущества: (вспом. задача эквивалентна выпуклой)

- Уходит из максимумов и седловых точек.
- На выпуклых функциях сходится как $O(1/k^2)$.
- Локальная квадратичная сходимость.

4.2. Ускоренная кубическая регуляризация

Задача: выпуклая функция, липшицев гессиан.

Инициализация: Выберем $x_0 \in E$, $M = 2L$ и $N = 12L$.
Вычислим $x_1 = T_L(x_0)$ и положим $\psi_1(x) = f(x_1) + \frac{N}{6}\|x - x_0\|^3$.
Итерация k, $(k \geq 1)$:

1. Вычислим $v_k = \arg\min\limits_{x \in E} \psi_k(x)$ и положим $y_k = \frac{k}{k+3} x_k + \frac{3}{k+3} v_k$.
2. Вычислим $x_{k+1} = T_M(y_k)$ и пересчитаем
   $\psi_{k+1}(x) = \psi_k(x) + \frac{(k+1)(k+2)}{2} \cdot [f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1} \rangle]$

Теорема: $\quad f(x_k) - f(x^*) \ \leq \ \frac{14\, L\, \|x_0 - x^*\|^3}{k(k+1)(k+2)}$.

NB:

- Используется аппарат оценочных функций (ср. с БГМ).
- Накапливаемая модель очень простая.

4.3. Модифицированный метод Гаусса-Ньютона

Задача: решить систему $F(x) = 0, \quad x \in E_1$.

Стандартный подход:

- Выбрать *функцию близости* $\phi(u)$ (вып., ≥ 0, Лип.=1).
- Решать задачу $\min\limits_{x \in E_1} \{ \, f(x) \equiv \phi(F(x)) \, \}$ с помощью *локальной модели* $\psi(x; y) = \phi\left(F(x) + \nabla F(x)(y - x)\right)$.

Получаем метод *Гаусса-Ньютона*: $x_{k+1} \in \operatorname{Arg} \min\limits_{y \in E_1} \psi(x_k; y)$

(нет глобальных оценок).

Заметим: если $\|\nabla F(x) - \nabla F(y)\| \leq M \|x - y\|$, то

$$f(y) \leq \psi(x, y) + \tfrac{1}{2} M \|y - x\|^2.$$

Метод: $\quad x_{k+1} \in \operatorname{Arg} \min\limits_{y \in E_1} \left[\psi(x; y) + \tfrac{1}{2} M \|y - x\|^2 \right]$.

Преимущества: (вспомогательная задача выпукла)

- Монотонное убывание целевой функции.
- Глобальные оценки скорости сходимости.
- Локальная квадратичная сходимость.

5.1. Сглаживание для явной модели целевой функции

Оценки Теории Сложности: (неулучшаемые)

- Негладкие задачи: $O(L_0(f)^2/\epsilon^2)$.
- Гладкие задачи: $O(L_1(f)^{1/2}/\epsilon^{1/2})$.

Слишком большая разница!

Модель: $f(x) = \max\limits_{u \in Q_2} \langle Ax, u \rangle_2 \quad \to \quad \min\limits_{x \in Q_1}.$

Сглаживание: $f_\mu(x) = \max\limits_{u \in Q_2} \{ \langle Ax, u \rangle_2 - \mu d_2(u) \}$, $\mu \geq 0$.

(Обозначим через $u_\mu(x)$ решение этой задачи.)

Теорема: $f(x) \geq f_\mu(x) \geq f(x) - \mu D_2$, $L_1(f_\mu) = \|A\|^2/\mu$.

Стратегия минимизации:

- Выбираем $\mu \approx \epsilon$ и применяем БГМ к f_μ.

Получаем схему с оценкой сложности $O(L_0(f)/\epsilon)$.

5.2. Условие обратного зазора

Сопряженная задача: $\min\limits_{u \in Q_2} \left[\phi(u) \stackrel{\mathrm{def}}{=} \max\limits_{x \in Q_1} \langle Ax, u \rangle_2 \right]$.

NB: $\phi(u) \leq f(x)$, но $f_{\mu_2}(x) \leq f(x)$ и $\phi(u) \leq \phi_{\mu_1}(u)$.

Возможно обеспечить выполнение условия *обратного зазора*:

$$(**): \qquad f_{\mu_2}(\bar{x}) \;\leq\; \phi_{\mu_1}(\bar{u}).$$

Тогда для $D_1 = \max\limits_{x \in Q_1} d_1(x)$, и $D_2 = \max\limits_{u \in Q_2} d_2(u)$ имеем

$$0 \leq \max\{f(\bar{x}) - f^*, f^* - \phi(\bar{u})\} \leq f(\bar{x}) - \phi(\bar{u}) \leq \mu_1 D_1 + \mu_2 D_2,$$

Пусть $T_{\mu_2}(x) = \arg\min\limits_{y \in Q_1} \left\{ \langle \nabla f_{\mu_2}(x), y - x \rangle_1 + \frac{L_1(f_{\mu_2})}{2} \|y - x\|_1^2 \right\}$.

Теорема: Пусть $\bar{x} \in Q_1$ и $\bar{u} \in Q_2$ удовлетворяют (**).
Выберем $\mu_1^+ = (1-\tau)\mu_1$, и $\hat{x} = (1-\tau)\bar{x} + \tau x_{\mu_1}(\bar{u})$,

$$\bar{u}_+ \;=\; (1-\tau)\bar{u} + \tau u_{\mu_2}(\hat{x}), \quad \bar{x}_+ = T_{\mu_2}(\hat{x}).$$

Тогда для $(\bar{x}_+, \bar{u}_+)$ верно (**) с μ_1^+ и μ_2 при τ: $\frac{\tau^2}{1-\tau} \leq \frac{\mu_1}{L_1(f_{\mu_2})}$.

NB: Переключающаяся стратегия дает $O(1/\epsilon)$.

5.3. Сглаживание в полуопределенной оптимизации

Как сглаживать симметричные функции собственных значений?

Пусть функция $f(\tau)$ задана рядом $f(\tau) = a_0 + \sum\limits_{k=1}^{\infty} a_k \tau^k$ с $a_k \geq 0$ при $k \geq 2$.

Для симметрической матрицы X рассмотрим следующую симметричную функцию собственных значений:

$$F(X) = \sum_{i=1}^{n} f(\lambda^{(i)}(X)).$$

Теорема: Для любого $X \in \operatorname{dom} F$ и $H \in \mathcal{S}_n$ имеем

$$\langle \nabla^2 F(X)H, H\rangle \leq \sum_{i=1}^{n} \nabla^2 f(\lambda^{(i)}(|X|))(\lambda^{(i)}(|H|))^2.$$

Примеры:

1. Квадрат матричной l_p-нормы:
$F_p(X) = \frac{1}{2}\|\lambda(X)\|^2_{(2p)} = \frac{1}{2}\langle X^{2p}, I_n\rangle_M^{1/p}$.
Тогда для любого матричного направления H имеем:

$$\langle \nabla^2 F_p(X)H, H\rangle_M \le (2p-1)\|\lambda(H)\|^2_{(2p)} = (2p-1)\|H\|^2_{(2p)}.$$

2. Энтропийное сглаживание максимального собственного значения. Рассмотрим функцию
$E(X) = \ln \sum\limits_{i=1}^{n} e^{\lambda^{(i)}(X)}$. Тогда при любом H

$$\langle \nabla^2 E(X)H, H\rangle_M \le \left[\sum_{i=1}^{n} e^{\lambda^{(i)}(X)}\right]^{-1} \sum_{i=1}^{n} e^{\lambda^{(i)}(X)} (\lambda^{(i)}(|H|))^2 \le \|H\|^2_{(\infty)}.$$

Применения: $\lambda_{\max}(A(x)$, $\rho(A(x))$.

Оценка трудоемкости соответствующих методов имеет порядок $O(1/\epsilon)$.

6.1. Относительная точность: однородная модель

Задача: Найти $f^* = \min\limits_x \{f(x) : \ x \in \mathcal{L}\}$, где

- $\mathcal{L} = \{x \in R^n : \ Cx = b\}$, C - это $p \times n$-матрица, и $b \neq 0$,
- выпуклая функция $f(x)$ является однородной.

Предположение: $B_{\|\cdot\|^*}(\gamma_0) \subseteq \partial f(0) \subseteq B_{\|\cdot\|^*}(\gamma_1)$.
Обозначим $\alpha = \frac{\gamma_0}{\gamma_1} < 1$, $\quad \|x_0\| = \min\limits_x \{\|x\| : \ Cx = b\}$.

Теорема: $\alpha f(x_0) \leq \gamma_0 \cdot \|x_0\| \leq f^* \leq f(x_0) \leq \gamma_1 \cdot \|x_0\|$.

Для $k := 0 \dots N$ **повторяем**
Compute $f(x_k)$ и $g(x_k)$.
$x_{k+1} := \pi_{\mathcal{L}} \left(x_k - \frac{R}{\sqrt{N+1}} \cdot \frac{g(x_k)}{\|g(x_k)\|^*} \right)$.
Результат: $G_N(R) = \arg\min\{f(x) : \ x = x_0, \dots, x_N\}$.

Теорема. Выберем $\hat{\rho} \stackrel{\mathrm{def}}{=} \frac{1}{\gamma_0} f(x_0)$, $N = \left\lfloor \frac{1}{\alpha^4 \delta^2} \right\rfloor$. Тогда
$f(G_N(\hat{\rho})) \leq (1 + \delta) \cdot f^*$.

Обновление

Положим $\hat{N} = \left\lfloor \frac{e}{\alpha^2} \cdot \left(1 + \frac{1}{\delta}\right)^2 \right\rfloor$.

Пусть $\hat{x}_0 = x_0$. Для $t \geq 1$ повторяем следующие операции:

$$\boxed{\begin{array}{l} \hat{x}_t := G_{\hat{N}}\left(\frac{1}{\gamma_0} f(\hat{x}_{t-1})\right); \\ \textbf{если } f(\hat{x}_t) \geq \frac{1}{\sqrt{e}} f(\hat{x}_{t-1}) \textbf{ то } T := t \textbf{ и СТОП}. \end{array}}$$

Теорема: $T \leq 1 + 2\ln\frac{1}{\alpha}$. При этом $f(\hat{x}_T) \leq (1+\delta) f^*$.
Общее число градиентных шагов не превосходит

$$\frac{e}{\alpha^2} \cdot \left(1 + \frac{1}{\delta}\right)^2 \cdot \left(1 + 2\ln\frac{1}{\alpha}\right).$$

NB: Для модели $F(Ax)$ применяя сглаживание к F можно добиться оценки $O(\frac{1}{\alpha\delta})$.

6.2. Эллипсоидальная аппроксимация выпуклых тел

NB: хорошая аппроксимация $\partial f(0)$ снижает трудоемкость.

Для $G \succ 0$ обозначим $W_r(v, G) = \{s \in E^* : \; \|s - v\|_G^* \leq r\}$.
β-аппроксимация тела C : $W_1(v, G) \subseteq C \subseteq W_\beta(v, G)$.
Параметр β называется *радиусом* аппроксимации.

Метод эллипсоидов (Хачиян): ($O(n)$ итераций)

- Центрально-симметричные тела: $\beta = O(n^{1/2})$.
- Множества общего вида: $\beta = O(n)$.
- Знако-инвариантные тела: $\beta = O(n^{1/2})$. (Диаг. матрицы).

Полученная метрика используется в методах нахождения решения с относительной точностью.

Пример: билинейная матричная игра с неотрицательными коэффициентами решается за

$$4e(1 + \ln(2\sqrt{n}))\sqrt{2n \ln m}\left(1 + \frac{1}{\delta}\right) \text{ итераций.}$$

Заключение

Правильное использование информации из Черного Ящика повышает эффективность методов:

- Можно формировать решения сопряженных задач.
- Строятся точные критерии остановки.
- Решение задач с относительной точностью.
- Увеличение скорости сходимости за пределы границ Теории Сложности.

Публикации

1. Yu. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming* (A), **103** (1), 127-152 (2005).

2. Yu. Nesterov. Excessive gap technique in non-smooth convex minimizarion. SIAM J. Optim. **16** (1), 235-249 (2005).

3. Yu. Nesterov. Smoothing technique and its applications in semidefinite optimization. *Mathematical Programming*, **110**(2), 245-259 (2007).

4. Yu. Nesterov. Dual extrapolation and its application for solving variational inequalities and related problems. *Mathematical Programming*, **109**(2-3), 319-344 (2007).

5. Yu. Nesterov. Modified Gauss-Newton scheme with worst-case guarantees for its global performance. *Optimization Methods and Software*, **22**(3) 469-483 (2007)

6. Yu. Nesterov. Rounding of convex sets and efficient gradient methods for linear programming problems. *Optimization Methods and Software*, **23**(1), 109-128 (2008).

7. Yu. Nesterov. Accelerating the cubic regularization of Newton's method on convex problems. *Mathematical Programming*, **112**(1) 159-181 (2008)

8. Yu. Nesterov. "Unconstrained Convex Minimization in Relative Scale". *Mathematics of Operations Research*, **34**(1), 180-193 (February 2009)

9. Yu. Nesterov. Primal-dual subgradient methods for convex problems. *Mathematical Programming*, **120**(1), 261-283 (2009).

10 Yu. Nesterov, L. Scrimali. Solving strongly monotone variational and quasi-variational inequalities. *Discrete and Continuous Dynamical Systems*, **31**(4), 1383-1296 (2011).

11. Yu. Nesterov. Barrier subgradient method. *Mathematical Programming*, **127**(1): 31-56 (2011).

12. Yu. Nesterov. Gradient methods for minimizing composite функцияs. *Mathematical Programming*, **140**(1), 125-161 (2013).

13. Yu. Nesterov, B. Polyak. Cubic regularization of Newton's method and its global performance. *Mathematical Programming*, **108**(1), 177-205 (2006).