Lecture 11.

Gradient methods for minimizing composite objective function

– Motivation.

– Composite Gradient Mapping.

– Gradient Method.

– Accelerated scheme.

– Computational experiments.

Climbing out from the Black Box

Problem:

$$\min_{x \in Q} f(x),$$

- Q is a closed convex set in R^n,
- f is a continuous convex function.

Black-box complexity for finding ϵ-solution:

- f is Lipschitz-continuous: $O\left(\frac{1}{\epsilon^2}\right)$ calls of oracle.
- ∇f is Lipschitz-continuous: $O\left(\frac{1}{\epsilon^{1/2}}\right)$ calls of oracle.
- f is L.-continuous, $n < \infty$: $O\left(n \ln \frac{1}{\epsilon}\right)$ calls of oracle.

Structural optimization:

1. *Self-concordant functions:* f is linear, Q is given by a special ν-s.c.barrier; $O\left(\sqrt{\nu} \ln \frac{1}{\epsilon}\right)$ iterations.

2. *Smoothing technique*: $f(x) = \max\limits_{u \in V} [\langle Ax, u \rangle - \phi(u)]$

with *simple* V and Q; $O\left(\frac{1}{\epsilon}\right)$ iterations.

Working horse: *Optimal method for smooth convex functions*

Problem: $\min\limits_{x \in Q} f(x)$ with $f \in C^{1,1}(Q)$, L_f is known.

Gradient mapping: (Nemirovsky, Yudin 1976)

$$T(x) = \arg\min_{y \in Q} \left\{ \langle \nabla f(x), y - x \rangle + \tfrac{L_f}{2} \|y - x\|^2 \right\}.$$

(The norm is Euclidean: $\|u\|^2 = \langle Bu, u \rangle$ with $B \succ 0$.)

Method: **for** $k \geq 0$ **do** (N. 1983, 2004, 2005)

1. Compute $f(x_k), \nabla f(x_k)$.

2. $y_k = T(x_k)$.

3. $v_k = \arg\min\limits_{x \in Q} \left\{ \frac{L_f}{2} \|x - x_0\|^2 + \sum\limits_{i=0}^{k} \frac{i+1}{2} \langle \nabla f(x_i), x \rangle \right\}.$

4. $x_{k+1} = \frac{2}{k+3} v_k + \frac{k+1}{k+3} y_k$.

Convergence:

$$f(y_k) - f(x^*) \leq \frac{2L_f \|x^* - x_0\|^2}{(k+1)^2},$$

where x^* is the optimal solution.

Main drawback: uses the upper bound for L_f.

Smoothing technique: $L_f = \frac{1}{\epsilon} \cdot \|A\|^2 \cdot R^2$,
where R is the size of V.

Our goals:

1. Develop special methods for a new class of structured convex problems.

2. Develop adjustable strategies for all class parameters.

Composite objective function

Problem:

$$\min_{x \in Q} \left[\, \phi(x) \stackrel{\mathrm{def}}{=} f(x) + \Psi(x) \, \right],$$

where Q is a closed convex set,

- f is differentiable, convex, and defined by BB-oracle.
- Ψ is a general closed convex function, which is *simple*.

Examples:

- $Q \equiv R^n$, $\quad \Psi(x) = \sigma_V(x) \stackrel{\mathrm{def}}{=} \begin{cases} 0, & \text{if } x \in V, \\ +\infty, & \text{otherwise.} \end{cases}$
- Denote $f^* = \min_{x \in Q} f(x)$. Let F be a ν-scb for Q, and

  $$\Psi(x) \stackrel{\mathrm{def}}{=} \tfrac{\epsilon}{\nu} F(x).$$

  Then, for $\hat{x} \in \operatorname{int} Q$, we have

  $$f(\hat{x}) \le f^* + \|\nabla \phi(\hat{x})\|^* \cdot \|\hat{x} - x^*\| + \epsilon.$$

- *Sparse least squares*:

  $$\phi(x) = \tfrac{1}{2}\|Ax - b\|_2^2 + \|x\|_1 \stackrel{\mathrm{def}}{=} f(x) + \Psi(x),$$

  where $\|\cdot\|_k$ denotes the standard l_k-norm.

Note: ϕ itself has no good properties.

Direct Help

Composite gradient mapping (S.Wright 2007)

For any $y \in Q$ and $L > 0$ define

$$\begin{aligned} m_L(y;x) &= f(y) + \langle \nabla f(y), x - y \rangle + \tfrac{L}{2}\|x-y\|^2 \\ &+ \boxed{\quad \Psi(x) \quad}, \qquad (\Psi \text{ is simple!}) \\ T \equiv T_L(y) &= \arg\min_{x \in Q} \; m_L(y;x), \\ g_L(y) &= L \cdot B(y - T_L(y)). \end{aligned}$$

Optimality conditions:

$$\langle \nabla f(y) + LB(T-y) + \xi_T, x - T \rangle \; \geq \; 0, \quad \forall x \in Q,$$

where $\xi_T \in \partial \Psi(T)$. In what follows, we denote

$$\phi'(T) \; = \; \underbrace{\nabla f(T) + \xi_T}_{\text{the right element}} \; \in \; \partial \phi(T).$$

Problem class: $\quad (\; \|s\|_* = \langle s, B^{-1}s \rangle^{1/2} \;)$

$$\|\nabla f(x) - \nabla f(y)\|_* \; \leq \; L_f \|x - y\|, \quad x, y \in Q,$$

Hence, for all $x, y \in Q$ we have

$$|f(x) - f(y) - \langle \nabla f(y), x - y \rangle| \; \leq \; \tfrac{L_f}{2}\|x-y\|^2.$$

Main properties

At any $y \in Q$, we have

$$\begin{aligned} \phi(y) - \phi(T_L(y)) &\geq \tfrac{2L-L_f}{2L^2} \|g_L(y)\|_*^2, \\ \langle \phi'(T_L(y)), y - T_L(y) \rangle &\geq \tfrac{L-L_f}{L^2} \|g_L(y)\|_*^2. \end{aligned}$$

Moreover, for $L \geq L_f$,

$$\phi(y) - \tfrac{1}{2L} \|g_L(y)\|_*^2 \geq m_L(y; T_L(y)),$$

and

$$m_L(y; T_L(y)) \geq \phi(T_L(y)).$$

We are ready to analyze the Gradient Method.

Let us fix $\gamma_u > 1$, $\gamma_d \geq 1$. Define

Gradient Iteration $\mathcal{G}(x, M)$

SET:	$L := M.$
REPEAT:	$T := T_L(x),$
	if $\phi(T) > m_L(x; T)$ **then** $L := L \cdot \gamma_u,$
UNTIL:	$\phi(T) \leq m_L(x; T).$ (Full relaxation)
OUTPUT:	$\mathcal{G}(x, M).T \; = \; T, \quad \mathcal{G}(x, M).L \; = \; L.$

Choose $L_0 \in (0, L_f]$ and a starting point $y_0 \in Q$.

Gradient Method $\mathcal{GM}(y_0, L_0)$

$$
\begin{array}{l}
y_{k+1} \; = \; \mathcal{G}(y_k, L_k).T, \;\; M_k \; = \; \mathcal{G}(y_k, L_k).L, \\
L_{k+1} \; = \; \max\{L_0, M_k/\gamma_d\}.
\end{array}
$$

Performance guarantees

1. Let N_k be the # of calls of oracle after k iterations. Then

$$N_k \;\leq\; \left[1 + \frac{\ln \gamma_d}{\ln \gamma_u}\right] \cdot (k+1) + \frac{1}{\ln \gamma_u} \cdot \left(\ln \frac{\gamma_u L_f}{\gamma_d L_0}\right)_+ .$$

For example,

$$\gamma_u \;=\; \gamma_d \;=\; 2 \;\Rightarrow\; \begin{cases} N_k \;\leq\; 2(k+1) + \log_2 \frac{L_f}{L_0}, \\ L_k \;\leq\; L_f. \end{cases}$$

2. For convex ϕ, we have

$$\phi(y_k) - \phi(x^*) \;\leq\; \frac{2\gamma_u L_f R^2}{k+2}.$$

3. Let ϕ be strongly convex on Q with parameter μ_ϕ.

If $\frac{\mu_\phi}{L_f} \geq 2\gamma_u$, then

$$\begin{aligned} \phi(y_k) - \phi(x^*) \;&\leq\; \left(\frac{\gamma_u L_f}{\mu_\phi}\right)^k (\phi(y_0) - \phi(y^*)) \\ &\leq\; \frac{1}{2^k}(\phi(y_0) - \phi(y^*)). \end{aligned}$$

Otherwise,

$$\phi(y_k) - \phi(x^*) \;\leq\; \left(1 - \frac{\mu_\phi}{4\gamma_u L_f}\right)^k \cdot (\phi(y_0) - \phi(y^*)).$$

Note: We can have $\mu_\phi \geq L_f$.

Estimate sequences

Problem:

$$\min_{\boxed{x \in R^n}} \; [\, \phi(x) \stackrel{\mathrm{def}}{=} f(x) + \Psi(x) \,], \quad \text{where}$$

- function f is convex with $L_f < \infty$,
- function Ψ is closed and strongly convex on E with known parameter $\mu_\Psi \geq 0$. (It may be $\mathrm{dom}\, \Psi \neq R^n$.)

We update recursively the following sequences.

- A minimizing sequence $\{x_k\}_{k=0}^{\infty}$.
- A sequence of increasing scaling coefficients $\{A_k\}_{k=0}^{\infty}$:

$$A_0 = 0, \quad A_k \stackrel{\mathrm{def}}{=} A_{k-1} + a_k, \quad k \geq 1.$$

- Sequence of estimate functions $(k \geq 0)$

$$\psi_k(x) \;=\; l_k(x) + \boxed{A_k \Psi(x)} + \tfrac{1}{2}\|x - x_0\|^2,$$

  where $x_0 \in \mathrm{dom}\, \Psi$ is a starting point, and $l_k(x)$ are linear functions.

New: we update the estimates for L_f, using

- the initial guess $L_0 \leq L_f$,
- two adjustment parameters $\gamma_u > 1$ and $\gamma_d \geq 1$.

Main relations: (Maintained recursively for $k \geq 0$)

$$\mathcal{R}_k^1: \; A_k \phi(x_k) \; \leq \; \psi_k^* \; \equiv \; \min_x \psi_k(x),$$

$$\mathcal{R}_k^2: \quad \psi_k(x) \; \leq \; A_k \phi(x) + \tfrac{1}{2}\|x - x_0\|^2, \; \forall x \in E.$$

Useful consequences:

- Taking $x = x^*$, we get

$$\phi(x_k) - \phi(x^*) \; \leq \; \tfrac{\|x^* - x_0\|^2}{2A_k}, \quad k \geq 1.$$

- Denoting $v_k = \arg\min_{x \in E} \; \psi_k(x)$, we obtain

$$\|x^* - v_k\| \; \leq \; \|x^* - x_0\|, \quad k \geq 1.$$

Dual Gradient Method: Let $\psi_0(x) = \frac{1}{2}\|x - v_0\|^2$.

$$y_k \; = \; \mathcal{G}(v_k, L_k).T, \quad M_k \; = \; \mathcal{G}(v_k, L_k).L,$$

$$L_{k+1} \; = \; \max\{L_0, M_k/\gamma_d\}, \quad a_{k+1} \; = \; \tfrac{1}{M_k},$$

$$\psi_{k+1}(x) \; = \; \psi_k(x) + \tfrac{1}{M_k}[f(v_k) + \langle \nabla f(v_k), x - v_k \rangle$$

$$+ \Psi(x)].$$

Justification:

Define $x_0 = y_0$, and $\phi_k \stackrel{\rm def}{=} \min\limits_{0 \le i \le k} \phi(y_i)$, $k \ge 1$. Then

$$\begin{aligned}
\psi^*_{k+1} &= \min_x \{ \psi_k(x) + \tfrac{1}{M_k}[f(v_k) + \langle \nabla f(v_k), x - v_k \rangle \\
&\quad + \Psi(x)]\} \quad \left(\psi_k(x) \ge \psi^*_k + \tfrac{1}{2}\|x - v_k\|^2 \right) \\
&\stackrel{\mathcal{R}^1_k}{\ge} A_k \phi_k + \min_x \{ \tfrac{1}{2}\|x - v_k\|^2 + \tfrac{1}{M_k}[f(v_k) \\
&\quad + \langle \nabla f(v_k), x - v_k \rangle + \Psi(x)]\} \\
&= A_k \phi_k + a_{k+1} m_{M_k}(v_k; y_k) \\
&\ge A_k \phi_k + a_{k+1} \phi(y_k) \ \ge \ A_{k+1} \phi_{k+1}.
\end{aligned}$$

Since $M_k \le \gamma_u L_f$, we get

$$\phi_k - \phi(x^*) \ \le \ \tfrac{\gamma_u L_f}{2k} \|x^* - v_0\|^2, \quad k \ge 1.$$

We can do better!

Accelerated method $\mathcal{A}(x_0, L_0, \mu)$ $\quad (\mu \le \mu_\Psi)$

DEFINE: $\psi_0(x) = \frac{1}{2}\|x - x_0\|^2$, $A_0 = 0$.

Iteration $k \ge 0$

SET: $L := L_k$.

REPEAT: Find a from $\frac{a^2}{A_k + a} = \frac{1 + \mu A_k}{L}$.

Set $y = \frac{A_k x_k + a v_k}{A_k + a}$, and compute $T_L(y)$.

if $\langle \phi'(T_L(y)), y - T_L(y) \rangle < \frac{1}{2L} \|\phi'(T_L(y))\|_*^2$,

then $L := L \cdot \gamma_u$. (Damped relaxation)

UNTIL: $\langle \phi'(T_L(y)), y - T_L(y) \rangle \ge \frac{1}{2L} \|\phi'(T_L(y))\|_*^2$.

DEFINE: $y_k := y,\ M_k := L,\ a_{k+1} := a$,

$L_{k+1} := M_k / \gamma_d,\ x_{k+1} := T_L(y_k)$,

$\psi_{k+1}(x) := \psi_k(x) + a_{k+1}[f(x_{k+1})$

$+ \langle \nabla f(x_{k+1}), x - x_{k+1} \rangle + \Psi(x)]$.

Performance guarantees

1. Let N_k be the # of calls of oracle after k iterations. Then

$$N_k \le 2 \cdot \left[1 + \frac{\ln \gamma_d}{\ln \gamma_u}\right] \cdot (k+1) + \frac{1}{\ln \gamma_u} \cdot \ln \frac{2\gamma_u L_f}{\gamma_d L_0}.$$

For example,

$$\gamma_u = \gamma_d = 2 \Rightarrow \begin{cases} N_k \le 4(k+1) + \log_2 \frac{L_f}{L_0} + 1, \\ L_k \le L_f. \end{cases}$$

2. For convex ϕ, we have

$$\phi(x_k) - \phi(x^*) \le \frac{4\gamma_u L_f \|x^* - x_0\|^2}{k^2}.$$

3. If Ψ is strongly convex with parameter μ_Ψ, then

$$\phi(x_k) - \phi(x^*) \le \gamma_u L_f \|x^* - x_0\|^2 \left[1 + \sqrt{\frac{\mu_\Psi}{8\gamma_u L_f}}\right]^{-2(k-1)}.$$

Note:

- We can have $\mu_\Psi \ge L_f$.

Computational experiments

<u>*Sparse least squares:*</u> for $\tau > 0$ compute

$$\phi_\tau^* \stackrel{\mathrm{def}}{=} \min_{x \in R^n} \left[\phi_\tau(x) \stackrel{\mathrm{def}}{=} \tfrac{1}{2}\|Ax - b\|_2^2 + \tau \|x\|_1 \right],$$

where $A \equiv (a_1, \ldots, a_n) \in R^{m \times n}$ with $m < n$. Note that

$$\begin{aligned}
&\min_{x \in R^n} \left[\tfrac{1}{2}\|Ax - b\|_2^2 + \tau\|x\|_1 \right] \\
&= \min_{x \in R^n} \max_{u \in R^m} \left[\langle u, b - Ax \rangle - \tfrac{1}{2}\|u\|_2^2 + \tau\|x\|_1 \right] \\
&= \max_{u \in R^m} \min_{x \in R^n} \left[\langle b, u \rangle - \tfrac{1}{2}\|u\|_2^2 - \langle A^T u, x \rangle + \tau\|x\|_1 \right] \\
&= \max_{u \in R^m} \left[\langle b, u \rangle - \tfrac{1}{2}\|u\|_2^2 : \ \|A^T u\|_\infty \le \tau \right] \\
&= \max_{u \in R^m} \left[\tau\langle b, v \rangle - \tfrac{1}{2}\tau^2\|v\|_2^2 : \ \|A^T v\|_\infty \le 1 \right]
\end{aligned}$$

Dual problem: project vector $\frac{b}{\tau}$ onto the dual polytop

$$\mathcal{D} = \{ y \in R^m : \ \|A^T y\|_\infty \le 1 \}.$$

We take $\tau = 1$ and terminate by ϕ^*. Other parameters:

$$\gamma_u = \gamma_d = 2, \ x_0 = 0, \ L_0 = \max_{1 \le i \le n} \|a_i\|_2^2 \le L_f, \ \mu = 0.$$

Denote by m_* the number of nonzero components in x^*.

$$\begin{array}{lcl} \text{Gap} & = & (\phi(x_k) - \phi(x^*))/(\phi(x_0) - \phi(x^*)), \\ \text{SpeedUp} & = & \text{Actual residual /Theoretical guarantee.} \end{array}$$

Problem 1: $n = 4000,\ m = 1000,\ m_* = 100.$

	PG			DG			AC		
Gap	K	#Ax	SpeedUp	K	#Ax	SpeedUp	K	#Ax	SpeedUp
1	1	4	0.21%	1	4	0.85%	1	4	0.04%
2^{-1}	3	8	0.20%	3	12	0.81%	5	36	0.49%
2^{-2}	10	29	0.24%	8	38	0.89%	10	76	1.01%
2^{-3}	28	83	0.32%	25	123	1.17%	19	148	1.90%
2^{-4}	159	476	0.88%	156	777	3.45%	55	436	8.24%
2^{-5}	557	1670	1.53%	565	2824	6.21%	103	820	14.28%
2^{-6}	954	2862	1.31%	941	4702	5.17%	138	1100	12.80%
2^{-7}	1255	3765	0.86%	1257	6282	3.45%	169	1348	9.58%
2^{-8}	1430	4291	0.49%	1466	7328	2.01%	199	1588	6.63%
2^{-9}	1547	4641	0.26%	1613	8080	2.13%	227	1812	4.35%
2^{-10}	1640	4920	0.14%	1743	8713	0.61%	253	2020	2.71%
2^{-11}	1722	5167	0.07%	1849	9243	0.33%	277	2212	1.61%
2^{-12}	1788	5364	0.04%	1935	9672	0.17%	301	2404	0.96%
2^{-13}	1847	5539	0.02%	2003	10013	0.09%	324	2588	0.55%
2^{-14}	1898	5693	0.01%	2061	10303	0.05%	348	2780	0.31%
2^{-15}	1944	5831	0.01%	2113	10563	0.05%	374	2988	0.19%
2^{-16}	1987	5961	0.00%	2164	10817	0.03%	404	3228	0.11%
2^{-17}	2029	6085	0.00%	2217	11083	0.02%	433	3460	0.06%
2^{-18}	2072	6215	0.00%	2272	11357	0.01%	484	3868	0.04%
2^{-19}	2120	6359	0.00%	2331	11652	0.00%	514	4108	0.02%
2^{-20}	2165	6495	0.00%	2448	12238	0.00%	521	4152	0.01%

Note: $2^{-20} \approx 10^{-6}$.

Problem 2: $n = 5000,\ m = 500,\ m_* = 100.$

	PG			DG			AC		
Gap	κ	#Ax	SpeedUp	κ	#Ax	SpeedUp	κ	#Ax	SpeedUp
1	1	4	0.24%	1	4	0.96%	1	4	0.04%
2^{-1}	2	6	0.20%	2	8	0.81%	3	20	0.23%
2^{-2}	5	17	0.21%	5	24	0.81%	7	56	0.66%
2^{-3}	11	33	0.19%	11	45	0.77%	11	88	0.85%
2^{-4}	38	113	0.30%	38	190	1.21%	26	208	2.54%
2^{-5}	234	703	0.91%	238	1189	3.69%	68	544	8.84%
2^{-6}	1027	3081	1.98%	1026	5128	7.89%	143	1144	19.51%
2^{-7}	2402	7206	2.31%	2387	11933	9.17%	215	1716	22.14%
2^{-8}	3681	11043	1.77%	3664	18318	7.05%	276	2208	18.23%
2^{-9}	4677	14030	1.12%	4664	23318	4.49%	333	2664	13.29%
2^{-10}	5410	16230	0.65%	5392	26958	2.61%	392	3132	9.15%
2^{-11}	5938	17815	0.36%	5879	29393	1.41%	447	3572	5.94%
2^{-12}	6335	19006	0.19%	6218	31088	0.77%	495	3956	3.67%
2^{-13}	6637	19911	0.10%	6471	32353	0.41%	538	4300	2.14%
2^{-14}	6859	20577	0.05%	6670	33348	0.21%	573	4580	1.23%
2^{-15}	7021	21062	0.03%	6835	34173	0.13%	606	4844	0.68%
2^{-16}	7161	21483	0.01%	6978	34888	0.05%	630	5036	0.37%
2^{-17}	7281	21842	0.01%	7108	35539	0.05%	655	5236	0.20%
2^{-18}	7372	22115	0.00%	7225	36123	0.03%	682	5452	0.11%
2^{-19}	7438	22313	0.00%	7335	36673	0.02%	701	5596	0.06%
2^{-20}	7492	22474	0.00%	7433	37163	0.01%	710	5668	0.05%

Primal-dual methods

Denote

$$\phi_*(u) \;=\; \tfrac{1}{2}\|u\|_2^2 - \langle b, u\rangle.$$

Then,

$$\phi(x) + \phi_*(u) \;\geq\; 0 \quad \forall x \in R^n,\; u \in \mathcal{D}, \qquad (1)$$

and equality is achieved for the primal-dual solution.

Lemma. Let

$$A_k \phi(x_k) \;\leq\; \min_{x \in R^n} \Big\{ \sum_{i=1}^{k} a_i [f(z_i) + \langle \nabla f(z_i), x - z_i\rangle]$$

$$+ A_k \Psi(x) + \tfrac{1}{2}\|x - z_0\|^2 \Big\}. \qquad (*)$$

Denote $u_i = b - Az_i$, $\bar{u}_k \;=\; \frac{1}{A_k} \sum_{i=1}^{k} a_i u_i$. Then

$$\phi(x_k) + \phi_*(\bar{u}_k) \;\leq\; 0.$$

Moreover, if B is diagonal, from $(*)$ we derive

$$\rho(\bar{u}_k) \;\stackrel{\text{def}}{=}\; \left[\sum_{i=1}^{n} (\; |\langle a_i, \bar{u}_k\rangle| - 1 \;)_+^2 \right]^{1/2} \;\leq\; \tfrac{2}{A_k}\|x^*\|.$$

Thus, $\rho(\cdot)$ is the *Dual Infeasibility Measure.*

$$\text{GAP} = \rho(u_k)/\rho(u_0), \quad \Delta\Phi = \phi(x_k) - \phi(x^*).$$

Problem 3: $n = 500,\ m = 50,\ m_* = 25.$

	DG				AC			
GAP	K	#AX	$\Delta\phi$	SpeedUp	K	#AX	$\Delta\phi$	SpeedUp
1	2	8	$2.5 \cdot 10^{0}$	8.26%	2	12	$3.1 \cdot 10^{0}$	0.56%
2^{-1}	5	25	$1.4 \cdot 10^{0}$	9.35%	10	80	$7.5 \cdot 10^{-1}$	6.51%
2^{-2}	13	64	$6.0 \cdot 10^{-1}$	13.17%	16	124	$4.7 \cdot 10^{-1}$	8.3%
2^{-3}	26	130	$3.9 \cdot 10^{-1}$	12.69%	21	164	$3.7 \cdot 10^{-1}$	7.41%
2^{-4}	48	239	$2.7 \cdot 10^{-1}$	12.32%	31	244	$2.6 \cdot 10^{-1}$	8.28%
2^{-5}	103	514	$1.6 \cdot 10^{-1}$	13.28%	49	388	$1.4 \cdot 10^{-1}$	10.2%
2^{-6}	243	1212	$8.3 \cdot 10^{-2}$	15.64%	73	580	$7.6 \cdot 10^{-2}$	11.51%
2^{-7}	804	4019	$3.0 \cdot 10^{-2}$	25.93%	110	876	$3.1 \cdot 10^{-2}$	13.21%
2^{-8}	1637	8183	$6.3 \cdot 10^{-3}$	26.41%	157	1252	$1.0 \cdot 10^{-2}$	13.3%
2^{-9}	3298	16488	$4.6 \cdot 10^{-4}$	26.6%	214	1704	$2.3 \cdot 10^{-3}$	12.47%
2^{-10}	4837	24176	$1.8 \cdot 10^{-7}$	19.33%	277	2208	$1.6 \cdot 10^{-4}$	10.36%
2^{-11}	4942	24702	$1.2 \cdot 10^{-14}$	9.97%	338	2688	$7.5 \cdot 10^{-5}$	7.7%
2^{-12}	5149	25734	$-1.3 \cdot 10^{-15}$	5.16%	445	3548	$1.9 \cdot 10^{-6}$	6.73%
2^{-13}	5790	28944	$-1.3 \cdot 10^{-15}$	2.92%	522	4160	$4.6 \cdot 10^{-8}$	4.64%
2^{-14}	6474	32364	0.0	2.67%	538	4280	$1.4 \cdot 10^{-6}$	4.04%

Note: $2^{-14} \approx 10^{-4}$.

Problem 4: $n = 1000,\ m = 100,\ m_* = 50.$

	DG				AC			
Gap	κ	#Ax	$\Delta\phi$	SpeedUp	κ	#Ax	$\Delta\phi$	SpeedUp
1	2	8	$3.7 \cdot 10^{0}$	6.41%	2	12	$4.7 \cdot 10^{0}$	0.5%
2^{-1}	5	24	$2.0 \cdot 10^{0}$	7.75%	8	60	$1.4 \cdot 10^{0}$	3.88%
2^{-2}	15	74	$1.0 \cdot 10^{0}$	11.56%	16	128	$8.8 \cdot 10^{-1}$	7.35%
2^{-3}	37	183	$6.9 \cdot 10^{-1}$	14.73%	23	184	$6.8 \cdot 10^{-1}$	7.94%
2^{-4}	83	414	$4.5 \cdot 10^{-1}$	16.49%	35	276	$4.6 \cdot 10^{-1}$	9.41%
2^{-5}	198	989	$2.4 \cdot 10^{-1}$	19.79%	54	428	$2.4 \cdot 10^{-1}$	11.11%
2^{-6}	445	2224	$7.8 \cdot 10^{-2}$	22.28%	81	644	$9.1 \cdot 10^{-2}$	12.59%
2^{-7}	1328	6639	$2.2 \cdot 10^{-2}$	33.25%	112	892	$3.5 \cdot 10^{-2}$	12.18%
2^{-8}	2675	13373	$4.1 \cdot 10^{-3}$	33.48%	157	1252	$1.0 \cdot 10^{-2}$	11.78%
2^{-9}	4508	22535	$5.6 \cdot 10^{-5}$	28.22%	217	1732	$2.6 \cdot 10^{-3}$	11.42%
2^{-10}	4702	23503	$2.7 \cdot 10^{-10}$	14.7%	295	2356	$3.1 \cdot 10^{-4}$	10.5%
2^{-11}	4869	24334	$-2.2 \cdot 10^{-15}$	7.61%	369	2944	$7.9 \cdot 10^{-5}$	8.25%
2^{-12}	6236	31175	$-2.2 \cdot 10^{-15}$	4.88%	469	3740	$7.8 \cdot 10^{-6}$	6.64%
2^{-13}	12828	64136	$-2.2 \cdot 10^{-15}$	5.02%	557	4440	$6.1 \cdot 10^{-7}$	4.69%
2^{-14}	16354	81766	$-4.4 \cdot 10^{-15}$	5.24%	580	4620	$2.2 \cdot 10^{-7}$	4.17%

Problem 1a: $n = 4000,\ m = 1000,\ m_* = 100.$

	DG				AC			
Gap	κ	#Ax	$\Delta\phi$	SpeedUp	κ	#Ax	$\Delta\phi$	SpeedUp
1	2	8	$2.3 \cdot 10^1$	2.88%	2	8	$2.3 \cdot 10^1$	0.25%
2^{-1}	5	24	$1.2 \cdot 10^1$	3.44%	9	68	$9.5 \cdot 10^0$	2.45%
2^{-2}	17	83	$5.8 \cdot 10^0$	6.00%	18	140	$4.6 \cdot 10^0$	4.80%
2^{-3}	44	219	$3.5 \cdot 10^0$	7.67%	27	212	$3.5 \cdot 10^0$	5.57%
2^{-4}	100	497	$2.7 \cdot 10^0$	8.94%	39	308	$2.9 \cdot 10^0$	5.79%
2^{-5}	234	1168	$1.9 \cdot 10^0$	10.51%	61	484	$2.1 \cdot 10^0$	7.15%
2^{-6}	631	3153	$1.0 \cdot 10^0$	14.18%	110	876	$9.9 \cdot 10^{-1}$	11.63%
2^{-7}	1914	9568	$1.0 \cdot 10^{-2}$	21.50%	170	1356	$2.7 \cdot 10^{-1}$	13.94%
2^{-8}	3704	18514	$4.6 \cdot 10^{-7}$	20.77%	235	1876	$5.7 \cdot 10^{-2}$	13.27%
2^{-9}	3731	18678	$1.4 \cdot 10^{-14}$	15.77%	325	2596	$4.2 \cdot 10^{-3}$	12.78%
2^{-10}	Line	search	failure ...		458	3660	$1.7 \cdot 10^{-4}$	12.68%
2^{-11}					592	4724	$1.7 \cdot 10^{-5}$	10.58%
2^{-12}					708	5652	$2.2 \cdot 10^{-6}$	7.58%
2^{-13}					813	6484	$1.8 \cdot 10^{-7}$	5.00%
2^{-14}					850	6784	$1.5 \cdot 10^{-7}$	4.47%

Reason for failure:

DG comes too close to the solution.