From Evolutionary Biology to Interior-Point Methods

Paul Tseng
Mathematics, University of Washington
Seattle

WCOM, UBC Kelowna
October 27, 2007

Joint work with Immanuel Bomze and Werner Schachinger (Univ. Vienna)

Talk Outline

- Affine-Scaling Method for Linearly Constrained Optimization

- Replicator Dynamics in Evolutionary Biology

- 1st-Order Interior-Point Methods for Linearly Constrained Optimization
  - ⋆ Convergence & Convergence Rate
  - ⋆ Numerical Tests

- Conclusions & Open Questions

Linearly Constrained Smooth Optimization

(P)
$$\boxed{\max_x \quad f(x) \quad \text{s.t.} \quad Ax = b,\ x \geq 0}$$

$f : \Re^n \to \Re$ is continuously diff., $A \in \Re^{m \times n}$ of rank m, $b \in \Re^m$

Seek a stationary pt: a feasible x ($Ax = b, x \geq 0$) with $x \perp \nabla f(x) - A^\top \lambda \leq 0$ for some $\lambda \in \Re^m$.

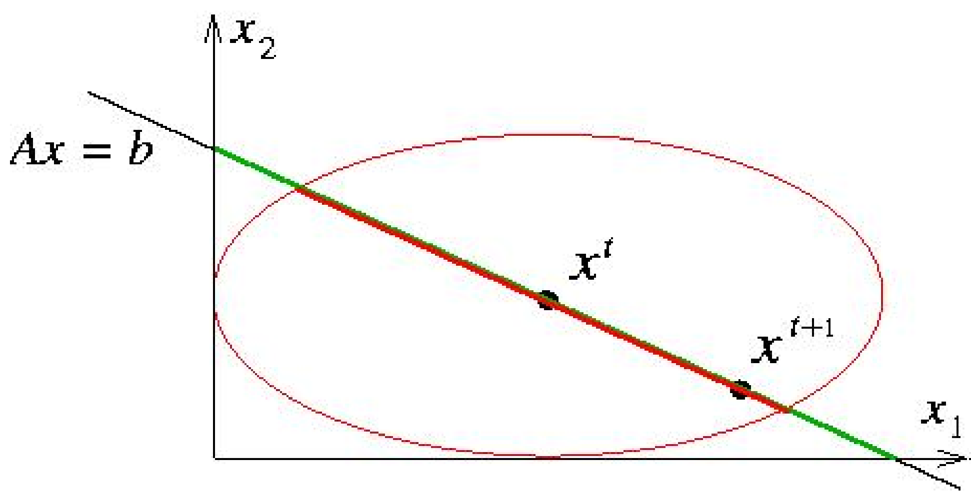
Primal Nondegeneracy: For any feasible x, the columns of A corresponding to $\{j \mid x_j \neq 0\}$ have rank m.

Affine-Scaling Method for (P)

Given $Ax^0 = b, x^0 > 0$, generate for $t = 0, 1, ...$

$$\boxed{\begin{array}{rcl} d^t & = & \arg\max_d \left\{ \nabla f(x^t)^\top d \mid Ad = 0,\ \|(X^t)^{-1} d\|_2 \le 1 \right\} \\ x^{t+1} & = & x^t + \alpha^t d^t > 0 \end{array}}$$

with $X^t = \mathrm{diag}(x^t)$, $\alpha^t > 0$ suitably chosen Dikin '67, '72



The AS method is simple, fairly efficient in practice, but difficult to analyze.

Barnes, Bonnans, Dikin, Gonzaga, Mascarenhas, Monma, Monteiro, Roos, Saigal, J. Sun, T, Tsuchiya, Vanderbei, Ye, ...

- When f is linear, $\{x^t\}$ and $\{f(x^t)\}$ converge linearly; Luo, T '92

the limit $\bar{x}$ attains maximum if α^t is not "too large"; Tsuchiya '91, Tsuchiya & Muramatsu '95

$\bar{x}$ may fail to attain maximum if α^t is "too large". Mascarenhas '97, Terlaky & Tsuchiya '99

- When f is concave or convex and assuming primal nondeg, every cluster pt of $\{x^t\}$ is a stationary pt of (P). Gonzaga & Carlos '90, Monteiro & Wang '98

What about general f? And convergence rate?

Replicator Dynamics in Evolutionary Biology

$$x_j^t \quad = \quad \text{fraction of } j\text{th genotype in pop. at time } t, \qquad j = 1, ..., n$$

Initially, $x_j^0 > 0$ and $\sum_j x_j^0 = 1$. $\quad x^t = (x_j^t)_{j=1}^n$ evolves according to

$$\boxed{x^{t+1} = \frac{X^t Q x^t}{(x^t)^\top Q x^t}, \qquad t = 0, 1, ...}$$

with adaptation coefficients $Q_{ii} > 0$ and $Q_{ij} \geq 0$ for all i, j ..., Haldane '32, ..., Moran '62, ...

- $\{x^t\}$ converges and its limit $\bar{x}$ satisfies $\bar{x}_j((Q\bar{x})_j - \bar{\lambda}) = 0$ for all j, with $\bar{\lambda} = \max_j (Q\bar{x})_j$; convergence rate is linear iff $(Q\bar{x})_j < \bar{\lambda}$ whenever $\bar{x}_j = 0$; otherwise $\|\bar{x} - x^t\| = O(1/\sqrt{t})$. Lyubich et al. '80

Connecting RD with AS

Rewrite RD iteration as

$$x^{t+1} - x^t = \frac{X^t[g^t - e\,(x^t)^\top g^t]}{(x^t)^\top g^t}$$

with $g^t = Qx^t$, $e = (1, ..., 1)^\top$.

Thus

$$x^{t+1} = x^t + \alpha^t d^t \quad \text{with} \quad d^t = X^t r(x^t)$$

where $\alpha^t = 1/(x^t)^\top g^t$ and

$$r(x) = \nabla f(x) - e\,x^\top \nabla f(x)$$

Solve for d^t in AS iteration when $A = e^\top$ yields

$$d^t \propto (X^t)^2 \left(g^t - e \, \frac{e^\top (X^t)^2 g^t}{\|x^t\|_2^2} \right)$$

with $g^t = \nabla f(x^t)$, $e = (1, ..., 1)^\top$.

Thus

$$x^{t+1} = x^t + \alpha^t d^t \quad \text{with} \quad d^t = (X^t)^2 r(x^t)$$

where $\alpha^t > 0$ and

$$r(x) = \nabla f(x) - e \, \frac{e^\top X^2 \nabla f(x)}{\|x\|^2}$$

1st-Order Interior-Point Methods for (P)

Given $Ax^0 = b, x^0 > 0$, generate for $t = 0, 1, \ldots$

$$\boxed{x^{t+1} = x^t + \alpha^t d^t \quad \text{with} \quad d^t = (X^t)^{2\gamma} r_\gamma(x^t)}$$

where $\gamma > 0$,

$$\boxed{r_\gamma(x) = \nabla f(x) - A^\top (AX^{2\gamma}A^\top)^{-1} AX^{2\gamma} \nabla f(x)}$$

and α^t is the largest $\alpha \in \{\alpha_0^t (\beta)^k\}_{k=0,1,\ldots}$ satisfying

$$f(x^t + \alpha d^t) \geq f(x^t) + \sigma \alpha (g^t)^\top d^t,$$

Armijo-type

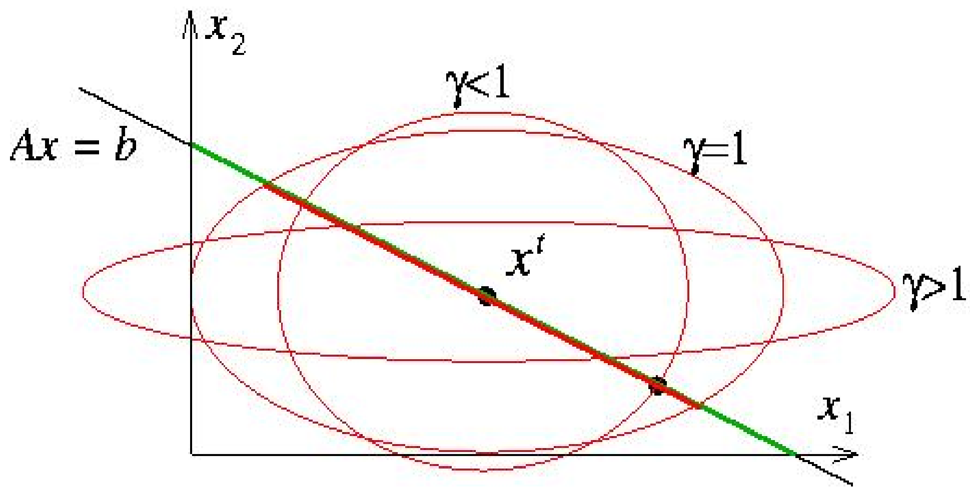
inexact LS

where $0 < \beta, \sigma < 1$, $g^t = \nabla f(x^t)$, and

$$0 < \alpha_0^t < \begin{cases} \infty & \text{if } d^t \geq 0; \\ \frac{-1}{\min_j d_j^t / x_j^t} & \text{else.} \end{cases}$$

$\gamma = 1/2 \quad \Longrightarrow \quad$ RD

$\gamma = 1 \quad \Longrightarrow \quad$ AS



This method is simple, suited for large problems ($n \geq 10000$).

Convergence of $\{x^t\}$? Convergence rate of $\{f(x^t)\}$, $\{x^t\}$?

Choosing γ?

Convergence Results: Assume $\{x \text{ feasible} \mid f(x) \geq f(x^0)\}$ is bounded. Then $x^t > 0$, $\{f(x^t)\} \uparrow$, and $\{x^t\}$, $\{d^t\}$ are bounded.

(a) Assume primal nondeg, f is concave or convex, and we choose $\inf_t \alpha_0^t > 0$, $\sup_t \alpha_0^t < \infty$. Then every cluster pt of $\{x^t\}$ is a stationary pt of (P).

(b) Assume f is quadratic and we choose $\inf_t \alpha_0^t > 0$. Then

$$v - f(x^t) = O\left(1/t^{1/\max\{\gamma, 2\gamma-1\}}\right),$$

with $v = \lim_{t\to\infty} f(x^t)$.

Assume in addition we choose $\gamma < 1$. Then $\{x^t\}$ converges and its limit $\bar{x}$ satisfies

$$\|\bar{x} - x^t\| = O\left(1/t^{\frac{1-\gamma}{2\gamma}}\right).$$

Under primal nondeg, $\bar{x}$ is a stationary pt of (P). Moreover, if $\gamma \leq \frac{1}{2}$ and $\bar{x} - r_\gamma(\bar{x}) > 0$, then $\{f(x^t)\}$ converges Q-linearly and $\{\|\bar{x} - x^t\|\}$ converges R-linearly. If instead $\sup_t \alpha_0^t < \infty$, $\gamma \geq \frac{1}{2}$ and $\bar{x} - r_\gamma(\bar{x}) \not> 0$, then $\{\|\bar{x} - x^t\|\}$ cannot converge linearly.

- Thus, $\gamma < 1$ seems preferrable.

Numerical Tests:

- Implement 1st-order IP method in Matlab. For Armijo LS, use $\beta = .5$, $\sigma = .1$,

$$\alpha_0^t = \min\left\{0.95\alpha_{\text{feas}}^t, \max\left\{10^{-5}, \frac{\alpha^{t-1}}{\beta^2}\right\}\right\}, \quad \alpha_{\text{feas}}^t = \frac{-1}{\min_j (d_j^t / x_j^t)},$$

with $\alpha^{-1} = \infty$.

- Numerical tests on

$$\max_x \quad f(x) \quad \text{s.t.} \quad e^\top x = 1,\ x \geq 0$$

with $-f$ from Moré-Garbow-Hillstrom set (least square), and $n = 1000$.

- Initial $x^0 = e/n$. Terminate when $\text{resid} := \|\min\{x^t, -r_\gamma(x^t)\}\| \leq tol$.

$f(x)$	γ	**#iter**	**#f-eval**	**cpu** (sec)	**obj**	**resid**
BAL	.8	†8	84	0.02	$9.98998 \cdot 10^{8}$	$1.4 \cdot 10^{-6}$
	1	7	8	0.03	$9.98998 \cdot 10^{8}$	$3.6 \cdot 10^{-8}$
	1.2	8	9	0.02	$9.98998 \cdot 10^{8}$	$3.7 \cdot 10^{-7}$
BT	.8	9146	27409	20.68	999.031	$9.9 \cdot 10^{-4}$
	1	17559	52665	23.31	999.055	$9.9 \cdot 10^{-4}$
	1.2	527757	$1.58 \cdot 10^{6}$	1192.55	999.081	$9.9 \cdot 10^{-4}$
DBV	.8	99	299	0.42	$4.9 \cdot 10^{-8}$	$9.8 \cdot 10^{-5}$
	1	146	440	0.52	$4.5 \cdot 10^{-8}$	$9.8 \cdot 10^{-5}$
	.2	240	722	1.02	$4.0 \cdot 10^{-8}$	$9.9 \cdot 10^{-5}$
EPS	.8	424	1269	1.89	$1.3 \cdot 10^{-6}$	$9.9 \cdot 10^{-4}$
	1	987	2958	3.52	$3.9 \cdot 10^{-6}$	$9.9 \cdot 10^{-4}$
	1.2	1963	5887	8.76	$8.0 \cdot 10^{-6}$	$9.4 \cdot 10^{-4}$
ER	.8	5	6	0.01	498.002	$5.2 \cdot 10^{-7}$
	1	7	8	0.03	498.002	$2.6 \cdot 10^{-7}$
	1.2	10	11	0.03	498.002	$3.7 \cdot 10^{-7}$
LR1	.8	20	21	0.04	$3.32834 \cdot 10^{8}$	$9.5 \cdot 10^{-7}$
	1	19	20	0.03	$3.32839 \cdot 10^{8}$	$3.5 \cdot 10^{-7}$
	1.2	20	21	0.03	$3.33481 \cdot 10^{8}$	$9.9 \cdot 10^{-7}$
VD	.8	22	74	0.04	$6.22504 \cdot 10^{22}$	$2.4 \cdot 10^{-9}$
	1	19	46	0.04	$6.22504 \cdot 10^{22}$	$2.6 \cdot 10^{-8}$
	1.2	18	50	0.05	$6.22504 \cdot 10^{22}$	$5.7 \cdot 10^{-9}$

Table 1: Performance of 1st-order IP method.

† Quit due to Armijo ascent condition not met when $\alpha < 10^{-20}$

Conclusions & Open questions

1. Theory and practice suggest $\gamma < 1$ is preferrable to $\gamma \geq 1$.

2. The method and its analysis readily extends to $0 \leq x \leq u$ by replacing X^t with $\min\{X^t, U - X^t\}$.

3. Convergence of $\{x^t\}$ when $\gamma \geq 1$ or when f is not quadratic?

4. Linear convergence of $\{f(x^t)\}$ and $\{x^t\}$ when $\gamma > \frac{1}{2}$ or when f is not quadratic?

5. Convergence of $\{x^t\}$ for 2nd-order AS method?

Convergence Proof Ideas

(a) $f(x^{t+1}) - f(x^t) \geq \sigma \alpha^t \|\eta^t\|^2$ with $\eta^t = (X^t)^\gamma r(x^t)$.

(b) If f is quadratic, use linearity of KT condition to show

$$\Delta^t := \upsilon - f(x^t) = O\left(\|\eta^t\|^{\min\{\frac{2}{1+\gamma},\frac{1}{\gamma}\}}\right).$$

(c) If f is quadratic and $\gamma < 1$, then

$$\|x^{t+1} - x^t\| = \alpha^t \|(X^t)^\gamma \eta^t\| = O(\|\eta^t\|) = O\left(\frac{\|\eta^t\|^2}{\|\eta^t\|}\right) = O\left(\frac{\Delta^t - \Delta^{t+1}}{(\Delta^t)^{\frac{1+\gamma}{2}}}\right) =$$
$$O\left(\int_{\Delta^{t+1}}^{\Delta^t} t^{-\frac{1+\gamma}{2}} dt\right) = O\left((\Delta^t)^{\frac{1-\gamma}{2}} - (\Delta^{t+1})^{\frac{1-\gamma}{2}}\right)$$

(d) Under primal nondegeneracy, r_γ is continuous on the feasible set.