Asymmetric Long-Step Primal-Dual IPM based on Dual Centering

Yurii Nesterov

Corvinus University of Budapest
and
SDS Chinese University of Hong Kong
and
Professor Emeritus at UCLouvain

June 5, 2025 (IMT, Lucca)

Workshop on Optimization for Learning and Control

Ref: arXiv:2503.10155v1 [math.OC]

Modern Optimization Challenges

Problems: Machine Learning, Data Mining, Artificial Intelligence.

New situation: Availability of big data arrays.

Consequences: Large-scale problems, treated by the

FIRST-ORDER METHODS

NB:

- In Machine Learning, we need to classify different data patterns.
- For that, we need to measure DISTANCES between the patterns.
- For that, we need to compute an appropriate METRIC.
- This can be done only by SEMIDEFINITE PROGRAMMING (LMI)

However: No activity observed in this direction up to now.

Reason: LMI have a reputation of DIFFICULT problems.

Our goal: reduce complexity of LMI up to that of Linear Programming.

Primal-dual pair of conic problems

Let $K \subset \mathbb{E}$ be a pointed closed convex cone, $\operatorname{int} K \neq \emptyset$ ($\equiv$ *proper*).
Then $K^* = \{s \in \mathbb{E}^* : \langle s, x \rangle \geq 0,\ x \in K\}$ is the *dual cone* (proper too).

$$f^* = \min_{\mathcal{F}_p = \left\{ \begin{array}{c} x \in K \\ Ax = b \end{array} \right.} \langle c, x \rangle = \max_{\mathcal{F}_d = \left\{ \begin{array}{c} y \in \mathbb{R}^m, s \in K^* \\ s + A^* y = c \end{array} \right.} \langle b, y \rangle,$$

Denote $\mathcal{F} = \mathcal{F}_p \times \mathcal{F}_d$. **Assumption:** $\boxed{\exists (\tilde{x}, \tilde{y}, \tilde{s}) \in \operatorname{rint} \mathcal{F}}$

Example 1: Let $f^* = \min\limits_{X \in \mathbb{S}^n_+} \Big\{ \langle C, X \rangle :\ \langle A_i, X \rangle = b^{(i)},\ i = 1, \dots, m \Big\}$.

Then $f^* = \max\limits_{y \in \mathbb{R}^m} \Big\{ \langle b, y \rangle :\ C \succeq A^*(y) = \sum\limits_{i=1}^m y^{(i)} A_i \Big\}$.

Log.-homogeneous SCB: $\boxed{F(\tau x) \equiv F(x) - \nu \ln \tau}$ $x \in \operatorname{int} K,\ \tau > 0$.

Dual SCB: $F_*(s) = \max\limits_{x \in \operatorname{int} K} \Big\{ - \langle s, x \rangle - F(x) \Big\}$. (*Regular Barriers*)

Primal-dual central path: denote $\mathcal{F}_0 = \operatorname{rint} \mathcal{F}$, and

$$z_t = (x_t, y_t, s_t) = \arg \min_{z \in \mathcal{F}_0} \Big\{ t[\langle c, x \rangle - \langle b, y \rangle] + F(x) + F_*(s) \Big\}.$$

Primal-Dual Functional Proximity Measure

Define $\boxed{\Omega(z) = F(x) + F_*(s) + \nu \ln \frac{\langle s,x \rangle}{\nu} + \nu}$

Theorem: for $z \in \mathcal{F}$, $\Omega(z) = 0$ iff $z = z_t$ and $t = \frac{\nu}{\langle s,x \rangle}$.

Predictor-corrector schemes

1. *Corrector Stage*.

We fix $t > 0$ and try to find a point $z \approx z(t)$ by minimizing $\Omega(z)$ with $\langle s, x \rangle = \langle c - A^* y, x \rangle = \langle c, x \rangle - \langle b, y \rangle = \frac{\nu}{t}$. This problem is *convex*.

2. *Predictor Stage (**general**)*.

If $z \approx z(t)$, we compute $d \approx z'(t)$ and move keeping

$$z + \alpha d \approx z(t + \alpha),\ \alpha > 0,$$

by checking $\Omega(z + \alpha d)$. Computation of $\Omega(z)$ is easy.

NB: For fast increase of $t > 0$, we need big α . By theory, we can take $t_+ = \left(1 + \frac{O(1)}{\lambda_\Phi(z(t))}\right) t$, with $\lambda_\Phi(z) = \langle \nabla \Phi(z), [\nabla^2 \Phi(z)]^{-1} \nabla \Phi(z) \rangle^{1/2}$.

For $\Phi = F + F_*$, we have $\boxed{\lambda_\Phi(z(t)) \equiv \nu^{1/2}}$

Symmetric Cones $\equiv$ Self-Scaled Barriers

Definition: $\boxed{F_*(\nabla^2 F(w)u) = F(x) - 2F(w) - \nu}$

(Covers LP, SDP, QP, see N.& Todd [97,98])

a). Find $z = (x, y, s) \approx z(t)$ and compute the *scaling point* $w \in \mathrm{int}\, K$:

$$\boxed{s = \nabla^2 F(w)x}$$

b). Compute *Affine-Scaling Direction* $d = (\Delta x, \Delta y, \Delta s)$ by the system

$$\Delta s + \nabla^2 F(w)\Delta x = -s, \quad A\Delta x \;=\; 0, \quad \Delta s + A^* \Delta y \;=\; 0.$$

Main advantage: Symmetry and convenient expression for $\Omega(z + \alpha d)$.

Example 2: Complexity of **a)**.

Let $K = \mathbb{S}^n_+$. Then $F(X) = -\ln\det X$ with $\nu = n$ and

$$\nabla F(X) \;=\; -X^{-1}, \quad \nabla^2 F(X)H \;=\; X^{-1}HX^{-1},$$

and $S = W^{-1}XW^{-1}$. Thus, $W^{-1} = X^{-1/2}[X^{1/2}SX^{1/2}]^{1/2}X^{-1/2}$. $\square$

Too difficult for normal engineers?

Primal and Dual Central Paths

Denote $g(x) = F(x)$, $x \in \mathrm{rint}\, \mathcal{F}_p$, the *restriction* of $F(\cdot)$ onto $\mathcal{F}_p$.

It is a SCB for $\mathcal{F}_p$ with $\nu_g \leq \nu$. Denote

$$x_t = \arg \min_{Ax=b} [\, f_t(x) \stackrel{\mathrm{def}}{=} t\langle c, x\rangle + F(x)\,].$$

For the dual problem, define

$$(y_t, s_t) = \arg \max_{s+A^*y=c} [\, \varphi_t(y,s) \stackrel{\mathrm{def}}{=} t\langle b, y\rangle - F_*(s)\,].$$

We can eliminate s and define the dual barrier function

$$\zeta(y) \stackrel{\mathrm{def}}{=} F_*(c - A^*y), \quad y \in \mathcal{F}_y \stackrel{\mathrm{def}}{=} \{y \in \mathbb{R}^m :\ c - A^*y \in K^*\}$$

$$\psi_t(y) \stackrel{\mathrm{def}}{=} \zeta(y) - t\langle b, y\rangle, \quad t > 0.$$

NB: $\zeta(\cdot)$ is an SCB for $\mathcal{F}_y$ with parameter $\nu_\zeta \leq \nu$.

Main Lemma 1. For any $t > 0$, we have

$$\boxed{\lambda_g^2(x_t) + \lambda_\zeta^2(y_t) = \nu}$$

Consequences

1. Primal and dual trajectories have *complementary complexities*.

2. No hope for a simultaneous big primal-dual step.

3. We get $\boxed{\nu \leq \nu_g + \nu_\zeta}$ If ν_ζ is small, then ν_g is big.

Lemma 2. Let set $\mathcal{F}_y$ be bounded. Then $\boxed{\nu_g = \nu \geq \nu_\zeta}$

Proof. There exists the analytic center y^*_ζ with $\nabla \zeta(y^*_\zeta) = 0$. Hence

$$\nu \;\geq\; \nu_g \geq \lim_{t\to 0} \lambda^2_g(x_t) \;=\; \lim_{t \to 0} [\nu - \lambda^2_\zeta(y_t)] \;=\; \nu. \quad \Box$$

Example 3. Consider $f^* = \min\limits_{X \succeq 0} \left\{ \langle I_n, X \rangle : \; Xa = b \right\}$ with $\langle a, b \rangle > 0$.

It has $\frac{n(n+1)}{2}$ variables and n linear equality constraints.
Denote $A_i = \frac{1}{2}(ae_i^T + e_i a^T)$. Then $\langle A_i, X \rangle = \; b^{(i)}, 1 \leq i \leq n$.

The dual problem is $\boxed{\max\limits_{y \in \mathbb{R}^n} \left\{ \langle b, y \rangle : \; I_n \succeq \frac{1}{2}(ya^T + ay^T) \right\}}$

$\zeta(y) = -\ln\det\left(I_n - \frac{1}{2}(ya^T + ay^T)\right)$. WLOG $a = \alpha e_1$ with $\alpha = \|a\|$.

$$\det S(y) = \left(1 - \tfrac{1}{2}\alpha y^{(1)}\right)^2 - \tfrac{1}{4}\alpha^2 \|y\|^2 \;=\; 1 - \alpha y^{(1)} - \tfrac{1}{4}\alpha^2 \sum_{i=2}^n (y^{(i)})^2.$$

Hence, $\boxed{\nu_\zeta = 1}$ and $\nu_g \geq n - 1$.

Dual Centering

Consider a dual centering process for $\boxed{\min\limits_{y \in \mathcal{F}_y} \left\{ \psi_t(y) = \zeta(y) - t\langle b, y\rangle \right\}}$

Newton Method: compute $\lambda_k = \langle \nabla\psi_t(y_k), [\nabla^2\zeta(y_k)]^{-1}\nabla\psi_t(y_k)\rangle^{1/2}$,

and update $\boxed{y_{k+1} = y_k - \frac{1}{1+\lambda_k}[\nabla^2\zeta_d(y_k)]^{-1}\nabla\psi_t(y_k), \quad k \geq 0}$

We get a point $\bar{y}$ with $\bar{\lambda} \stackrel{\text{def}}{=} \langle \nabla\psi_t(\bar{y}), [\nabla^2\zeta(\bar{y})]^{-1}\nabla\psi_t(\bar{y})\rangle^{1/2} \leq \beta \; < \; 1$.

For $\boxed{\bar{d} \stackrel{\text{def}}{=} [\nabla^2\zeta(\bar{y})]^{-1}\nabla\psi_t(\bar{y})}$ we have $\|\bar{d}\|_{\bar{y}} = \langle \nabla^2\zeta(\bar{y})\bar{d}, \bar{d}\rangle^{1/2} \; = \; \bar{\lambda}$.

Dual Gambit Rule Denote $\bar{s} = c - A^*\bar{y} \in \operatorname{int} K^*$.

$$\boxed{\begin{array}{rclrcl} \hat{y} & = & \bar{y} + \bar{d}, & \hat{s} & = & c - A^*\hat{y} \; \equiv \; \bar{s} - A^*\bar{d} \\ \bar{w}_* & = & \sqrt{t}\,\bar{s}, & \hat{x} & = & \nabla^2 F_*(\bar{w}_*)\hat{s}. \end{array}}$$

Example 4. For dual SDO-problem, $\zeta(y) = -\ln\det\left(C - \sum\limits_{i=1}^{m} y^{(i)}A_i\right)$.

Form $\hat{y} = \bar{y} + \bar{d}$ and $\hat{S} = C - \sum\limits_{i=1}^{m}\hat{y}^{(i)}A_i$. Then $\hat{X} = \frac{1}{t}\bar{S}^{-1}\hat{S}\bar{S}^{-1}$. □

Geometric interpretation

Denote $\bar{x} = -\frac{1}{t}\nabla F_*(\bar{s}) \in \mathrm{int}\, K$. Then $\bar{s} = -\frac{1}{t}\nabla F(\bar{x})$.

Theorem 1. We have $\boxed{A\hat{x} = b}$ and $\|\hat{x} - \bar{x}\|_{\bar{x}} = \|\hat{s} - \bar{s}\|_{\bar{s}} = \bar{\lambda} \leq \beta < 1$.

Hence, $\hat{x} \in \mathrm{int}\, K$, $\hat{s} \in \mathrm{int}\, K^*$, and $\hat{z} = (\hat{x}, \hat{y}, \hat{s}) \in \mathcal{F}_0$.

Lemma 3. We have $tc + \nabla F(\bar{x}) + \nabla^2 F(\bar{x})(\hat{x} - \bar{x}) = tA^*(\bar{y} - d)$.

Hence, $$\boxed{\hat{x} = \arg\min_{Ax=b} \left\{ \langle \nabla f_t(\bar{x}), x - \bar{x} \rangle + \frac{1}{2}\|x - \bar{x}\|^2_{\bar{x}} \right\}}$$

NB: point $\hat{z}$ is close to the primal-dual CP: $\boxed{\Omega(\hat{z}) \leq \omega_* \left(\frac{2\beta}{(1-\beta)^2} \right)}$
where $\omega_*(\tau) = -\tau - \ln(1 - \tau)$.

Affine-Scaling Direction

Define primal scaling point: $\bar{w} \stackrel{\mathrm{def}}{=} \sqrt{t}\bar{x} \in \mathrm{int}\, K$. Then $\boxed{\hat{s} = \nabla^2 F(\bar{w})\hat{x}}$

Define $\boxed{\Delta s + \nabla^2 F(\bar{w})\Delta x = -\hat{s}, \quad A\Delta x = 0, \quad \Delta s + A^*\Delta y = 0}$

where $\bar{x}$, $\bar{s}$, and $\hat{s}$ are defined by the Dual Gambit rule.

We use it for PREDICTOR STEP.

Functional Proximity Measure for SSB

Lemma 4. Let K be symmetric and for points $x, w \in \mathrm{int}\, K$, $s \in \mathrm{int}\, K^*$, we have $\boxed{s = \nabla^2 F(w) x}$ Define $\Delta z = (\Delta x, \Delta y, \Delta s)$ by equations

$$\Delta s + \nabla^2 F(w)\Delta x = -s, \quad A\Delta x = 0, \quad \Delta s + A^*\Delta y = 0.$$

Then

$$\xi_z(\alpha) \stackrel{\mathrm{def}}{=} \Omega(z + \alpha\Delta z) - \Omega(z)$$
$$\boxed{= F_*(s + \alpha\Delta s) + F_*\left(s - \frac{\alpha}{1-\alpha}\Delta s\right) - 2F_*(s)}$$

Corollary 1: $\boxed{\xi_{\hat{z}}(\alpha) = \zeta(\hat{y} + \alpha\Delta y) + \zeta\left(\hat{y} - \frac{\alpha}{1-\alpha}\Delta y\right) - 2\zeta(\hat{y})}$

Lemma 5. The norm of ASD is bounded as follows:

$$\|\Delta y\|_{\hat{y}} \le \frac{\beta + \lambda_\zeta(y)}{1-\beta} \le \hat{c}_\beta \stackrel{\mathrm{def}}{=} \frac{\beta + \sqrt{\nu_\zeta}}{1-\beta}.$$

Thus, for symmetric cones, we get $\alpha - \frac{2}{\sqrt{\nu}}\beta \ge \frac{1}{2\hat{c}_\beta + 1}\left(\frac{\delta}{1+\delta} - \frac{2\beta(3+\beta)}{1-\beta}\right)$.

This leads to $\boxed{O\left(\sqrt{\nu_\zeta} \ln \frac{1}{\epsilon}\right)}$ worst-case complexity bound for our

Long-Step Primal-Dual Method

NB: This is the first time we get something better than $\boxed{O\left(\sqrt{\nu} \ln \frac{1}{\epsilon}\right)}$.

Predictor-Corrector & Dual Gambit Rule (PCDG)

We associate both centering and predicting strategies with dual problem.

Initialization. Choose $\beta \in (0, \frac{1}{4}]$ and $A = \delta + \omega_* \left(\frac{2\beta}{(1-\beta)^2} \right)$ with $\delta > 0$.
Compute $y_0 \in \mathcal{F}_y$ with $\|\nabla \zeta(y_0)\|_{y_0} \le \frac{1}{2}\beta$ and set $t_0 = \frac{\beta}{2\|b\|_{y_0}}$.

***k*th iteration ($k \ge 0$)**

a) For $y_k \in \operatorname{int} \mathcal{F}_y$ and $t_k > 0$, form $\boxed{B_k = [\nabla^2 \zeta(y_k)]^{-1}}$, $g_k = \nabla \psi_{t_k}(y_k)$, $d_k = B_k g_k$, and $\lambda_k = \langle g_k, d_k \rangle^{1/2}$. Define $s_k = c - A^* y_k \in \operatorname{int} K^*$.

b) **If** $\lambda_k > \beta$, **Then** $y_{k+1} = y_k - \frac{d_k}{1+\lambda_k}$, $t_{k+1} = t_k$. (Newton Step)

c) **Else** (Predictor Step)

1. Set $\boxed{w_k^* = t_k^{1/2} s_k,\ \hat{y}_k = y_k + d_k,\ \hat{s}^k = c - A^* \hat{y}_k,\ \hat{x}_k = \nabla^2 F_*(w_k^*) \hat{s}_k}$
2. Find Affine-Scaling Direction $\Delta z_k = (\Delta x_k, \Delta y_k, \Delta s_k)$ by
$$\boxed{\Delta x_k + \nabla^2 F_*(w_k^*) \Delta s_k = -\hat{x}_k, \quad A \Delta x_k = 0, \quad \Delta s_k + A^* \Delta y_k = 0}$$
3. Compute the step $\alpha_k > 0$ from equation $\boxed{\Omega(\hat{z}_k + \alpha_k \Delta z_k) = A}$
4. Define $z_{k+1} = \hat{z}_k + \alpha_k \Delta z_k$ and $t_{k+1} = t(z_{k+1}) = \frac{\nu}{(1-\alpha_k)\langle \hat{s}_k, \hat{x}_k \rangle}$.

Easy Linear Matrix Inequalities

Lemma 6. Let matrix $C \in \mathbb{R}^{n\times n}$, $C \succ 0$, and $A \in \mathbb{R}^{m\times n}$, $m < n$, has full row rank. Then SCB $\boxed{\zeta(y) \stackrel{\mathrm{def}}{=} -\ln\det(C - A^T D(y) A)}$ with $D(y) = \mathrm{Diag}(y)$, has the following short representation

$$\boxed{\zeta(y) = -\ln\det(G^{-1} - D(y)) - \ln(\sigma\gamma)}$$

where $G = AC^{-1}A^T$, $\sigma = \det C$, and $\gamma = \det G$. For its domain

$$\mathrm{dom}\,\zeta(y) = \{y \in \mathbb{R}^m : \ D(y) \prec G^{-1}\},$$

we get the best possible value $\boxed{\nu_\zeta = m}$ **NB:** Interesting case $\boxed{n \to \infty}$

Corollary 2. Let $C \in \mathbb{R}^{n\times n}$, $C \succ 0$, and $A_i = A_i^T \in \mathbb{R}^{n\times n}$, $i = 1, \dots, m$.

If $\zeta(y) = -\ln\det\left(C - \sum\limits_{i=1}^m y^{(i)} A_i\right)$, then $\boxed{\nu_\zeta \le \min\left\{n, \sum\limits_{i=1}^m \mathrm{rank}(A_i)\right\}}$

Numerical Experiments

Nonsmooth Optimization Problem (*Low-Rank Quadratic Interpolation*):

$$\min_{X=X^T\in\mathbb{R}^{n\times n}} \Big\{ \sum_{i=1}^n |\lambda_i(X)| : \ \langle Xa_i, a_i\rangle = b_i, \quad i=1,\dots,m \Big\},$$

where $b\in\mathbb{R}^m$ and $A^T \stackrel{\mathrm{def}}{=} (a_1,\dots,a_m)$ has full rank $(G \stackrel{\mathrm{def}}{=} AA^T \succ 0)$.

SDO-formulation:

$$\min_{X_1,X_2\succeq 0} \Big\{ \langle I_n, X_1+X_2\rangle : \ \langle (X_1-X_2)a_i, a_i\rangle = b_i, \quad i=1,\dots,m \Big\}.$$

The dual problem is as follows:

$$\boxed{f_* = \max_{y\in\mathbb{R}^m} \Big\{ \langle b, y\rangle : \ -I_n \preceq A^T D(y) A \preceq I_n \Big\}}$$

Thus, we have to choose the following dual barrier function

$$\zeta(y) = -\ln\det(I_n - A^T D(y) A) - \ln\det(I_n + A^T D(y) A)$$

$$\boxed{= -\ln\det(G^{-1} - D(y)) - \ln\det(G^{-1} + D(y)) - 2\ln\sigma}$$

where $\sigma = \det G$. Since $\mathcal{F}_y$ is bounded, we have $\nabla^2\zeta(y) \succ 0$.

For Method PCDG, choose $y_0 = 0$. Then $\nu_g = 2n$ and $\boxed{\nu_\zeta = m}$

Results for Random Problems

Parameters: $\beta = 0.2$, $A = 2$, $\epsilon = 10^{-8}$, and $y_0 = 0 \in \mathbb{R}^m$.

Average number of Predictor Steps (100 problems)

$m \setminus n$	64	128	256	512	1024
32	$9.0 \pm 9.6\%$	$8.2 \pm 9.3\%$	7.1 ± 4.2	$6.9 \pm 3.5\%$	$6.6 \pm 7.4\%$
64		$9.9 \pm 7.8\%$	$7.8 \pm 5.5\%$	$7.1 \pm 3.4\%$	$6.9 \pm 3.9\%$
128			$9.9 \pm 6.2\%$	$7.9 \pm 4.3\%$	$7.0 \pm 2.0\%$
256				$9.5 \pm 5.7\%$	$7.9 \pm 3.7\%$
512					$9.5 \pm 5.3\%$

Average number of Corrector Steps (100 problems)

$m \setminus n$	64	128	256	512	1024
32	$31.9 \pm 13.6\%$	$29.0 \pm 13.9\%$	24.7 ± 7.3	$26.1 \pm 6.0\%$	$24.9 \pm 7.4\%$
64		$37.3 \pm 10.3\%$	$27.2 \pm 5.9\%$	$25.4 \pm 4.8\%$	$26.0 \pm 4.7\%$
128			$38.4 \pm 8.8\%$	$27.6 \pm 5.4\%$	$26.0 \pm 3.5\%$
256				$36.8 \pm 7.1\%$	$28.3 \pm 5.0\%$
512					$36.5 \pm 6.2\%$

Typical dynamics ($m = 256$, $n = 1024$)

$N_{\rm pred}/N_{\rm corr}$	$\langle \hat{s}_k, \hat{x}_k \rangle$	t_k	Bisections
0/0		31.25	
1/1	$9.7 \cdot 10^1$	$2.4 \cdot 10^2$	9
2/4	$8.6 \cdot 10^0$	$7.1 \cdot 10^2$	7
3/6	$2.9 \cdot 10^0$	$2.5 \cdot 10^3$	8
4/10	$8.4 \cdot 10^{-1}$	$1.3 \cdot 10^4$	8
5/16	$1.5 \cdot 10^{-1}$	$2.6 \cdot 10^5$	10
6/21	$7.9 \cdot 10^{-3}$	$6.8 \cdot 10^7$	13
7/26	$3.0 \cdot 10^{-5}$	$8.9 \cdot 10^{10}$	15
8/30	$6.8 \cdot 10^{-9}$	$3.2 \cdot 10^{11}$	

1. Our SDO-problems can be solved in the time comparable with the time required by Linear Optimization.

2. For implementation of PCDG-method, only the standard Cholesky factorization is needed.

3. Complexity of iteration in the Line Search (by bisections) is $O(m)$.

For competitors:

$$f_*^{-1} = \min_{y \in \mathbb{R}^m} \left\{ \sigma_{\max} \left(\sum_{i=1}^m y^{(i)} a_i a_i^T \right) : \ \langle b, y \rangle = 1 \right\}$$

Conclusion

1. We presented a new methodology for solving primal-dual problems of Conic Optimization by predictor-corrector schemes.

- We run the corrector process only in the dual space.
- Approximation of the dual path is used as a SCALING POINT for a new feasible primal-dual pair generated by the Dual Gambit Rule.
- Size of predictor step is defined by *Functional Proximity Measure*.
- The computational complexity is very reasonable even for SDO-problems. We need only the standard Cholesky factorization.

2. The main motivation is that in many problems the complexity of primal and dual problems are different.

3. We prove that the standard assumption on boundedness of the dual feasible set leads to $\nu_g = \nu \geq \nu_\zeta$.

4. An automatic switching to the best value of the barrier parameter is ensured by the Functional Proximity Measure.

5. Complexity of some SDO-problems is similar to Linear Optimization.

6. Numerical results are very promising (local superlinear convergence).

7. It is interesting to check performance of other methods on our test set.