KTH Matematik

5B1817 Tillämpad ickelinjär optimering

Föreläsning 11

Semidefinit programmering

Semidefinit programmering

- Semidefinit programmering är en utvidgning av ickelinjär programmering till att omfatta matrisolikheter.
- Vi tittar genomgående på linjär semidefinit programmering. Resonemanget kan generaliseras till konvex semidefinit programmering.
- Tillämpningar finns inom många områden, exempelvis reglerteknik, signalbehandling, matematisk systemteori och kombinatorisk optimering.
- Inrepunktsmetoder har visat sig mycket kraftfulla för att lösa semidefinita programmeringsproblem.

Linjärprogrammering

Till det *primala* linjärprogrammeringsproblemet

$$(PLP) \qquad \begin{array}{ll} \text{min} & c^T x \\ \text{då} & Ax = b, \\ & x \geq 0. \end{array}$$

hör det *duala* linjärprogrammeringsproblemet

$$(DLP) \qquad \begin{array}{ll} \text{max} & b^T y \\ \text{då} & A^T y + s = c, \\ & s \geq 0. \end{array}$$

Vi vet att *stark dualitet* gäller för linjärprogrammering, dvs (PLP) och (DLP) har samma optimalvärde (om båda är tillåtna).

Metrik för semidefinita matriser

För x och y i $\mathbb{R}^n$ ges inreprodukten av $x^T y = \sum_{j=1}^{n} x_j y_j$.

Den tillhörande normen är den euklidiska normen, $\|x\|_2 = \sqrt{x^T x}$.

Låt $\mathcal{S}^n$ beteckna rummet av symmetriska $n \times n$-matriser.

För X och Y i $\mathcal{S}^n$ ges inreprodukten av

$$\operatorname{trace}(X^T Y) = \operatorname{trace}(XY) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij} y_{ij},$$

där för en $n \times n$-matris M, $\operatorname{trace}(M) = \sum_{j=1}^{n} m_{jj}$.

Den tillhörande normen är frobeniusnormen, $\|X\|_F = \sqrt{\operatorname{trace}(X^T X)}$.

Metrik för semidefinita matriser, forts.

För en $n \times n$-matris M definierar vi trace$(M) = \sum_{j=1}^{n} m_{jj}$.

Det gäller dessutom att trace$(M) = \sum_{j=1}^{n} \eta_j(M)$, där $\eta_j(M)$ betecknar $j$te egenvärdet till M.

Vi kommer att använda båda dessa egenskaper av spåret. Vi kommer också använda trace$(AB) =$ trace(BA).

Påstående. *Om* $A \succeq 0$ *och* $B \succeq 0$ *tillhör* $\mathcal{S}^n$ *är* trace$(AB) \geq 0$*.*

Bevis. Om $A = A^T \succeq 0$ finns det L så att $A = LL^T$. Då gäller trace$(AB) =$ trace$(LL^TB) =$ trace$(L^TBL) \geq 0$. $\square$

Påstående. *Om* $A \succeq 0$ *och* $B \succeq 0$ *tillhör* $\mathcal{S}^n$ *är* trace$(AB) = 0$ *om och endast om* $AB = 0$*.*

Semidefinit programmering

I semidefinit programmering är X en matris i $\mathcal{S}^n$. Linjärprogrammering fås som specialfallet där X är diagonal.

I LP har vi bivillkor $A_i^T x = b_i$, $i = 1, \ldots, m$, där $A_i^T \in I\!R^n$ är rad i av bivillkorsmatrisen A.

I SDP kommer A_i tillhöra $\mathcal{S}^n$ för $i = 1, \ldots, m$, och bivillkoret blir trace$(A_i X) = b_i$, $i = 1, \ldots, m$.

På samma sätt har vi i LP målfunktion $c^T x$, där $c \in I\!R^n$.

I SDP blir målfunktionen trace(CX), där $C \in \mathcal{S}^n$.

Ett semidefinit programmeringsproblem på standardform

Ett semidefinit programmeringsproblem på standardform kan skrivas som

$$
(PSDP) \qquad \begin{array}{ll} \text{min} & \text{trace}(CX) \\ \text{då} & \text{trace}(A_i X) = b_i, \quad i = 1, \dots, m, \\ & X = X^T \succeq 0. \end{array}
$$

(Om C och A_i, $i = 1, \dots, m$, är diagonala kan X väljas diagonal och vi har (PLP).)

Precis som för LP är formen inte så intressant. Det väsentliga är att vi har ett linjärt problem med likheter och matrisolikheter.

$(PSDP)$ är ett konvext problem.

Duala problem för semidefinit programmering

Till det primala problemet

$$(PSDP) \qquad \begin{array}{ll} \min & \operatorname{trace}(CX) \\ \text{då} & \operatorname{trace}(A_iX) = b_i, \quad i = 1, \dots, m, \\ & X = X^T \succeq 0, \end{array}$$

hör ett dualt problem

$$(DSDP) \qquad \begin{array}{ll} \max & b^T y \\ \text{då} & \sum_{i=1}^m A_i y_i + S = C, \\ & S = S^T \succeq 0. \end{array}$$

Dualen kan härledas med hjälp av lagrangerelaxering.

Duala problem för semidefinit programmering

Om X tillåten till $(PSDP)$ och y, S är tillåten till $(DSDP)$ gäller att

$$\begin{aligned}\text{trace}(CX) - b^T y &= \text{trace}((\sum_{i=1}^m A_i y_i + S)X) - b^T y \\ &= \text{trace}(SX) + \sum_{i=1}^m (\text{trace}(A_i X) - b_i) y_i \\ &= \text{trace}(SX) \geq 0.\end{aligned}$$

Därmed gäller svag dualitet, dvs dualen är väldefinierad.

Stark dualitet gäller dock inte alltid, till skillnad från LP.

Exempelproblem

För givet $C \in \mathcal{S}^n$, betrakta följande semidefinita programmeringspar,

$$(PSDP) \qquad \begin{array}{ll} \text{min} & \text{trace}(CX) \\ \text{då} & \text{trace}(X) = 1, \quad X = X^T \succeq 0, \end{array}$$

respektive

$$(DSDP) \qquad \begin{array}{ll} \text{max} & y \\ \text{då} & Iy \preceq C. \end{array}$$

Då blir dualitetsgapet noll. Optimalvärdet, liksom y^*, är minsta egenvärdet till C. $X^* = uu^T$, där u är en egenvektor som svarar mot minsta egenvärdet till C, normerad så att $\|u\|_2 = 1$.

Stark dualitet gäller dock inte alltid för SDP.

Relativa inre

För att ge villkor för stark dualitet behöver vi definiera det *relativa inre* till de tillåtna områdena.

Låt $F(PSDP)$ beteckna tillåtna området till $(PSDP)$, dvs

$$F(PSDP) = \{X \in \mathcal{S}^n : \text{trace}(A_i X) = b_i,\ i = 1, \dots, m,\ X \succeq 0\}.$$

Det *relativa inre* till $(PSDP)$ definieras som

$$F^0(PSDP) = \{X \in F(PSDP) : X \succ 0\}.$$

Analogt,

$$F(DSDP) = \{(y, S) \in \mathbb{R}^m \times \mathcal{S}^n : \sum_{i=1}^{m} A_i y_i + S = C,\ S \succeq 0\},$$
$$F^0(DSDP) = \{(y, S) \in F(DSDP) : S \succ 0\}.$$

Icektomt relativt inre ger stark dualitet

Sats. *Om* $F^0(PSDP) \neq \emptyset$ *är dualitetsgapet noll och* $(DSDP)$ *har en optimallösning.*

Beviset kan ges med hjälp av separerande hyperplanssatsen.

Corollary. *Om* $F^0(DSDP) \neq \emptyset$ *är dualitetsgapet noll och* $(PSDP)$ *har en optimallösning.*

Corollary. *Om* $F^0(PSDP) \neq \emptyset$ *och* $F^0(DSDP) \neq \emptyset$ *så är dualitetsgapet noll och både* $(PSDP)$ *och* $(DSDP)$ *har optimallösningar.*

Omskrivning av ett semidefinit programmeringsproblem

Ett semidefinit programmeringsproblem på standardform kan ekvivalent skrivas som

$$
\begin{array}{ll}
\min & \operatorname{trace}(CX) \\
\text{då} & \operatorname{trace}(A_i X) = b_i, \quad i = 1, \dots, m, \\
& X = X^T, \\
& u^T X u \geq 0 \ \ \forall\, u \in I\!R^n : \|u\|_2 = 1.
\end{array}
$$

Vi får ett oändligt antal olikhetsbivillkor. Problemet blir därmed mer komplext än ett LP-problem.

Hur löser vi $(PSDP)$? Generalisering av simplexmetoden inte uppenbar. Inrepunktsmetoder generaliseras dock ganska rättframt.

En alternativ formulering av ett semidefinit programmeringsproblem

$$
(PSDP) \qquad \begin{array}{ll} \text{min} & \operatorname{trace}(CX) \\ \text{då} & \operatorname{trace}(A_i X) = b_i, \quad i = 1, \dots, m, \\ & X = X^T \succeq 0. \end{array}
$$

Vi kan ekvivalent formulera $(PSDP)$ som

$$
(PSDP') \qquad \begin{array}{ll} \text{min} & \operatorname{trace}(CX) \\ \text{då} & \operatorname{trace}(A_i X) = b_i, \quad i = 1, \dots, m, \\ & X = X^T, \\ & \eta_j(X) \geq 0, \quad j = 1, \dots, n, \end{array}
$$

där $\eta_j(X)$ betecknar $j$te egenvärdet till X.

En nackdel är att $(PSDP')$ är ickedifferentierbar.

En barriärtransformation för semidefinit programmeringsproblem

$$(PSDP_\mu) \quad \begin{array}{ll} \min & \text{trace}(CX) - \mu \sum_{j=1}^n \ln(\eta_j(X)) \\ \text{då} & \text{trace}(A_i X) = b_i, \quad i = 1, \dots, m, \\ & X = X^T \succ 0. \end{array}$$

Notera att $\sum_{j=1}^n \ln(\eta_j(X)) = \ln \prod_{j=1}^n \eta_j(X) = \ln \det(X)$, vilket ger

$$(PSDP_\mu) \quad \begin{array}{ll} \min & \text{trace}(CX) - \mu \ln \det(X) \\ \text{då} & \text{trace}(A_i X) = b_i, \quad i = 1, \dots, m, \\ & X = X^T \succ 0. \end{array}$$

Differentierbart problem.

Derivativan av barriärtermen

Vi kan skriva $X = \sum_{k=1}^{n} \sum_{l=k}^{n} x_{kl} E_{kl},$

där $E_{kk} = e_k e_k^T$ och $E_{kl} = e_k e_l^T + e_l e_k^T,\ k < l.$

Påstående. *Om $X \in \mathcal{S}^n$, $X \succ 0$, gäller* $\dfrac{\partial \ln \det(X)}{\partial x_{kl}} = \operatorname{trace}(X^{-1} E_{kl}).$

Bevis. Vi får $\ln\det(X + tE_{kl}) = \ln\det(X) + \ln\det(I + tX^{-1}E_{kl})$. Låt η_j, $j = 1, \ldots, n$, beteckna egenvärdena till $X^{-1}E_{kl}$. Då gäller att

$$\begin{aligned}\ln\det(I + tX^{-1}E_{kl}) &= \ln \prod_{j=1}^{n} (1 + t\,\eta_j) = \sum_{j=1}^{n} \ln(1 + t\,\eta_j) \\ &= t \sum_{j=1}^{n} \eta_j + o(t) = t \operatorname{trace}(X^{-1}E_{kl}) + o(t).\end{aligned}$$

Resultatet följer genom att låta $t \to 0$. $\square$

Optimalitetsvillkor för det semidefinita barriärproblemet

Optimalitetsvillkoren för $(PSDP_\mu)$ ges av

$$
\begin{aligned}
\text{trace}(CE_{kl}) - \mu\,\text{trace}(X^{-1}E_{kl}) - \sum_{i=1}^{n} \text{trace}(A_i E_{kl}) y_i = 0,\ 1 \le k \le l \le n, \\
\text{trace}(A_i X) = b_i, \quad i = 1, \dots, m, \\
X = X^T.
\end{aligned}
$$

Ekvivalent,

$$
\begin{aligned}
\sum_{i=1}^{n} A_i y_i + \mu X^{-1} &= C, \\
\text{trace}(A_i X) &= b_i, \quad i = 1, \dots, m, \\
X &= X^T.
\end{aligned}
$$

Primal-dual form av optimalitetsvillkoren

Låt $S = \mu X^{-1}$, vilket ger

$$\begin{aligned}\sum_{i=1}^{n} A_i y_i + S &= C,\\ \text{trace}(A_i X) &= b_i, \quad i = 1, \dots, m,\\ XS &= \mu I.\end{aligned}$$

Detta är det primal-duala ekvationssystemet som ger barriärtrajektorian, där vi dessutom implicit kräver $X \succ 0$ och $S \succ 0$.

Observera att ekvationssystemet ger $X \in \mathcal{S}^n$ och $S \in \mathcal{S}^n$ eftersom $A_i \in \mathcal{S}^n$, $i = 1, \dots, m$, och $C \in \mathcal{S}^n$.

Lösningen är primalt respektive dualt tillåten med skillnad i målfunktionsvärde som ges av $n\mu$.

Icketomt relative inre garanterar existens av barriärtrajektorian

Den primal-duala formen av ekvationen för barriärtrajektorian är

$$\sum_{i=1}^{n} A_i y_i + S = C,$$
$$\text{trace}(A_i X) = b_i, \quad i = 1, \ldots, m,$$
$$XS = \mu I, \quad X \succ 0, \quad S \succ 0.$$

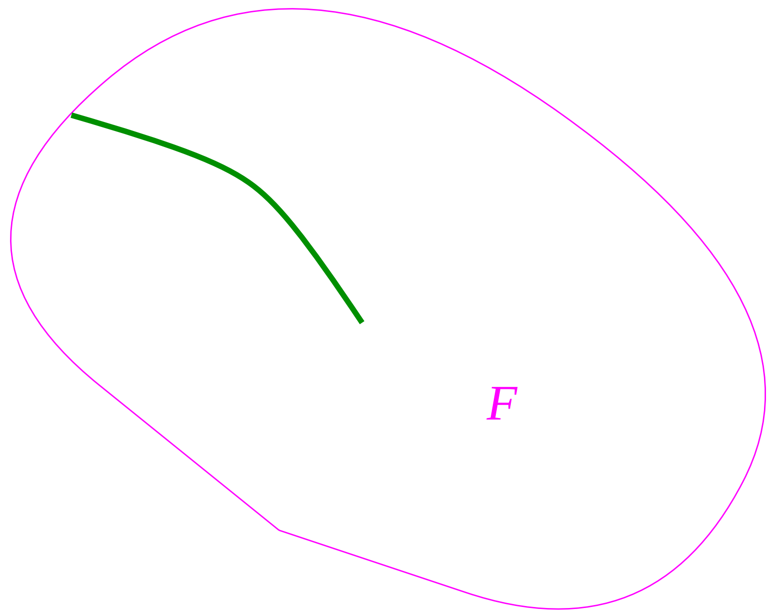
Sats. *Om $F^0(PSDP) \neq \emptyset$, $F^0(DSDP) \neq \emptyset$ och $\{A_i\}_{i=1}^m$ är linjärt oberoende så är barriärtrajektorian väldefinierad för $\mu > 0$.*

Detta kan visas genom att visa att för ett positivt μ är ovanstående ekvationssystem ekvivalenta med optimalitetsvillkoren till problemet

$$\min_{X \in F^0(PSDP)} \quad \text{trace}(CX) - \mu \ln \det X.$$

Barriärtrajektorian

Givet lämpliga villkor leder barriärtrajektorian till optimallösningen.



Inrepunktsmetoder följer trajektorian approximativt.

Primal-duala inrepunktsmetoder

Antag att X är tillåten till $(PSDP)$ med $X \succ 0$, och antag att y, S är tillåtna till $(DSDP)$ med $S \succ 0$.

Låt $\mu = \dfrac{\text{trace}(XS)}{n}$ vara uppskattningen av barriärparametern.

Låt $\sigma \in [0, 1]$ beteckna önskad reduktion av dualitetsgapet.

Det betyder att vi försöker lösa

$$\begin{aligned} \sum_{i=1}^{m} A_i y_i + S &= C, \\ \text{trace}(A_i X) &= b_i, \quad i = 1, \dots, m, \\ XS &= \sigma \mu I. \end{aligned}$$

En iteration i en primal-dual inrepunktsmetod

- Givet tillåtna X^k, y^k och S^k så att $X^k \succ 0$ and $S^k \succ 0$;
- $\mu^k \leftarrow \dfrac{\text{trace}(X^k S^k)}{n}$, $\sigma^k \in [0,1]$;
- Beräkna ΔX^k, Δy^k and ΔS^k;
- Beräkna steglängd α^k;
- $X^{k+1} \leftarrow X^k + \alpha^k \Delta X^k$, $y^{k+1} \leftarrow y^k + \alpha^k \Delta y^k$, $S^{k+1} \leftarrow S^k + \alpha^k \Delta S^k$;

Vi kräver $X^{k+1} \succ 0$ and $S^{k+1} \succ 0$.

Metoder skiljer sig i val av σ^k, val av steg och val av steglängd.

Rättframma Newtoniterationer

Sista blocket av ekvationer kan skrivas som $XS - \sigma\mu I = 0$, vilket ger Newtoniterationerna

$$
\begin{aligned}
\textstyle\sum_{i=1}^{m} A_i \Delta y_i + \Delta S &= 0, \\
\operatorname{trace}(A_i \Delta X) &= 0, \quad i = 1, \dots, m, \\
X \Delta S + \Delta X S &= \sigma\mu I - XS,
\end{aligned}
$$

eller som $SX - \sigma\mu I = 0$, vilket ger

$$
\begin{aligned}
\textstyle\sum_{i=1}^{m} A_i \Delta y_i + \Delta S &= 0, \\
\operatorname{trace}(A_i \Delta X) &= 0, \quad i = 1, \dots, m, \\
S \Delta X + \Delta S X &= \sigma\mu I - SX.
\end{aligned}
$$

I allmänhet är lösningarna olika, och de är inte symmetriska.

Reduktion av dualitetsgapet

Antag att ΔX, Δy, ΔS uppfyller

$$\begin{aligned}
\textstyle\sum_{i=1}^m A_i \Delta y_i + \Delta S &= 0, \\
\mathrm{trace}(A_i \Delta X) &= 0, \quad i = 1, \dots, m, \\
X \Delta S + \Delta X S &= \sigma \mu I - XS.
\end{aligned}$$

Då gäller att $\mathrm{trace}(\Delta X \Delta S) = -\sum_{i=1}^m \mathrm{trace}(A_i \Delta X) \Delta y_i = 0$.

Alltså,

$$\begin{aligned}
\mathrm{trace}((X + \alpha \Delta X)(S + \alpha \Delta S)) &= \mathrm{trace}(XS) + \alpha\, \mathrm{trace}(X \Delta S + \Delta X S) \\
&= (1 - \alpha(1 - \sigma))\, \mathrm{trace}(XS).
\end{aligned}$$

Samma relation för båda lineariseringarna.

Strategier för att välja σ

Relationen

$$\operatorname{trace}((X+\alpha \Delta X)(S+\alpha \Delta S)) = (1-\alpha(1-\sigma))\operatorname{trace}(XS)$$

implicerar att det är önskvärt att ha σ liten och α stor. Dessa mål strider i allmänhet mot varandra.

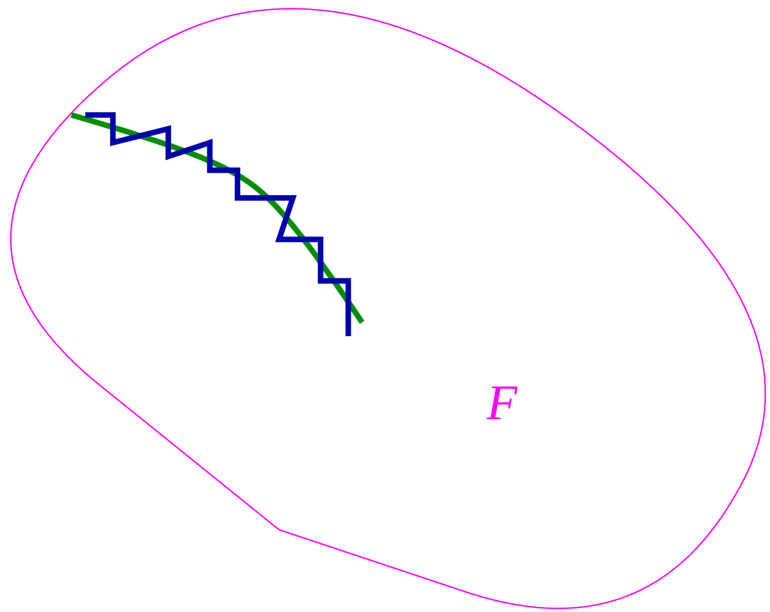
Tre huvudstrategier:

- Kort-stegsmetod, σ nära 1.
- Lång-stegsmetod, σ väsentligt mindre än 1.
- Prediktions-korrektionsmetod, $\sigma = 0$ varje jämn iteration och $\sigma = 1$ varje udda iteration.

Kort-stegsmetod

Vi kan välja $\sigma^k = 1 - \delta/\sqrt{n}$, $\alpha^k = 1$.

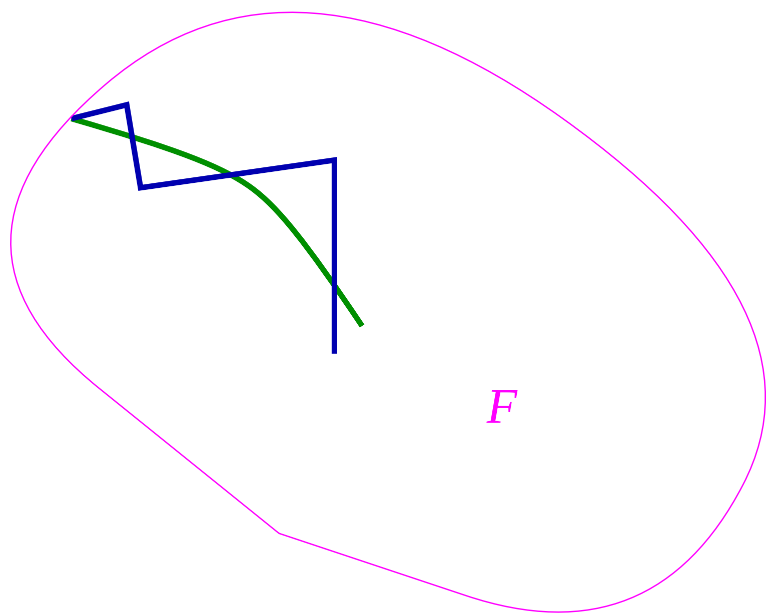
Iterationspunkterna stannar nära trajektorian.



Polynomiell komplexitet. Normalt inte tillräckligt effektiv i praktiken.

Lång-stegsmetod

Vi kan välja $\sigma^k = 0.1$, α^k given av närhet till trajektorian.

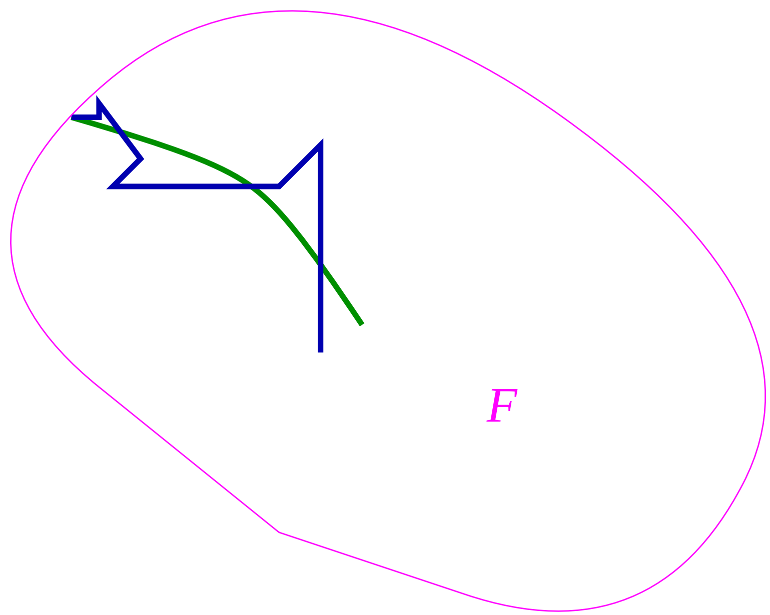


Polynomiell komplexitet.

Prediktions-korrektionsmetod

$\sigma^k = 0$, α^k given av närhet till trajektorian för k jämn.

$\sigma^k = 1$, $\alpha^k = 1$ för k udda.



Polynomiell komplexitet.

Symmetrisering av newtonekvationen

Det finns olika sätt att symmetrisera det sista blocket av ekvationer.

Ett exempel är Alizadeh-Haeberly-Overton-riktninger (AHO-riktningen), där

$$\frac{1}{2}XS + \frac{1}{2}SX - \sigma\mu I = 0.$$

Detta ger motsvarande block av newtonekvationen som

$$\frac{1}{2}X\Delta S + \frac{1}{2}\Delta XS + \frac{1}{2}S\Delta X + \frac{1}{2}\Delta SX = \sigma\mu I - \frac{1}{2}XS + \frac{1}{2}SX.$$

Lösningen ΔX, ΔS är nu symmetrisk.

Mer generellt kan vi införa $\mathcal{S}_P(U) = \frac{1}{2}\left(PUP^{-1} + (PUP^{-1})^T\right)$ för en ickesingulär $n \times n$-matris P och betrakta $\mathcal{S}_P(XS) = \sigma\mu I$.

Inrepunktsmetoder utan tillåten startpunkt

En kraftfull egenskap hos inrepunktsmetoder är att de inte behöver ha initialpunkter i det relativa inre. Det räcker att $X^0 \succ 0$ och $S^0 \succ 0$, men de linjära ekvationerna behöver inte vara uppfyllda.

Newtonekvationerna modifieras på motsvarande sätt.

$$
\begin{aligned}
\textstyle\sum_{i=1}^m A_i \Delta y_i + \Delta S &= C - \textstyle\sum_{i=1}^m A_i y_i - S, \\
\operatorname{trace}(A_i \Delta X) &= b_i - \operatorname{trace}(A_i X), \quad i = 1, \dots, m, \\
\mathcal{S}_P(X \Delta S + \Delta X S) &= \sigma \mu I - \mathcal{S}_P(XS).
\end{aligned}
$$

Om $\alpha = 1$, blir iterationspunkterna tillåtna.