- LECTURE 4 [illegible]

DUALITY THEORY IS A VALUABLE TOOL THAT CHARACHERIZES WHEN AND HOW WE CAN SOLVE THE ORIGINAL PROBLEM. IN SOME SETTINGS, WE ARE SATISFIED WITH AN APPROXIMATION TO THE ORIGINAL PROBLEM.

PENALTY METHODS

$$\begin{aligned} \min\ & f(x) \\ \text{s.t.}\ & g(x) \le 0 \\ & h(x) = 0 \end{aligned} \longrightarrow P_{count}(x) = \sum_i \big(g_i(x) > 0\big) + \sum_j \big(h_j(x) \neq 0\big)$$

AND WE INSTEAD SOLVE:

$$\min_x f(x) + \rho \cdot P_{count}(x) \quad \text{WITH INCREASING } \rho.$$

DRAWBACK: DISCONTINUOUS FUNCTION $P_{count}(x)$.

ALTERNATIVE: QUADRATIC PENALTIES.

$$P_{quad}(x) = \sum_i \max(g_i(x), 0)^2 + \sum_j h_j(x)^2$$

IS THERE A CONNECTION BETWEEN $P_{quad}(\cdot)$ AND AUGMENTED LAGRANGE MULTIPLIERS?

RECALL (FOR EQUALITY CONSTRAINTS):

$$P_{Lagrange}(x) = \frac{1}{2}\eta \cdot \sum_i (h_i(x))^2 + \sum_i \lambda_i \cdot h_i(x)$$

E.X. $h(x) = Ax - b$. THEN:

$$P_{Lagrange}(x) = \lambda^T (Ax - b) + \frac{\eta}{2} \|Ax - b\|_2^2$$

DRAWBACK OF PENALTY METHODS AND ADMM:

- ADMM (AND VARIANTS) HAVE BEEN TRADITIONALLY APPLIED TO LINEAR EQUALITY (MOSTLY) AND INEQUALITY CONSTRAINTS.
- PENALTY METHODS LIKE QUADRATIC ONES ARE LESS AGGRESSIVE WHEN CONSTRAINTS ARE VIOLATED.

• NEWTON'S METHOD

- WHAT IT IS:

$$x_{t+1} = x_t + \Delta x \quad \text{WHERE} \quad \Delta x = -\nabla^2 f(x_t)^{-1} \nabla f(x_t)$$

PSD OF $\nabla^2 f(x)$ IMPLIES THAT:

$$\nabla f(x_t)^T \Delta x = -\nabla f(x_t)^T \nabla^2 f(x_t)^{-1} \nabla f(x_t) < 0$$

I.E., NEWTON'S DESCENT IS A DESCENT DIRECTION

- INTERPRETATIONS OF NEWTON'S METHOD:
  - i) MINIMIZER OF SECOND-ORDER APPROXIMATION OF f.
  - ii) SINGLE-STEP MINIMIZER OF LINEAR EQUATIONS/ QUADRATIC FORMS

- KEY PROPERTY OF NEWTON'S METHOD: AFFINE INVARIANCE

SUPPOSE $T \in \mathbb{R}^{n \times n}$ IS NONSINGULAR, AND DEFINE: $\bar{f}(y) = f(Ty) = f(x)$

THEN:
$$\nabla \bar{f}(y) = T^T \cdot \nabla f(x)$$
$$\nabla^2 \bar{f}(y) = T^T \cdot \nabla^2 f(x) \cdot T$$

ASSUME WE PERFORM NEWTON'S STEPS ON $\bar{f}$ & y:

$$\Delta y = -\left(T^T \nabla^2 f(x) T\right)^{-1} \cdot T^T \cdot \nabla f(x)$$
$$= -T^{-1} \cdot \nabla^2 f(x)^{-1} \cdot \nabla f(x)$$
$$= T^{-1} \Delta x$$

THUS: $x + \Delta x = T \cdot (y + \Delta y)$

WHY IS THIS IMPORTANT?
- CHANGE OF VARIABLES CAN ABRUPTLY CHANGE THE CONDITION NUMBER
- GD SLOWS DOWN
- NEWTON IS NOT AFFECTED

- CONVERGENCE ANALYSIS.

ASSUMPTIONS: $mI \preceq \nabla^2 f(x) \preceq MI$

$$\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L \cdot \|x - y\|_2$$

OVERVIEW OF ANALYSIS:

IF $\|\nabla f(x_t)\|_2 \geq \eta$: $f(x_{t+1}) - f(x_t) \leq -\gamma$

IF $\|\nabla f(x_t)\|_2 < \eta$: $\frac{L}{2m^2} \|\nabla f(x_{t+1})\|_2 \leq \left(\frac{L}{2m^2} \cdot \|\nabla f(x_t)\|_2\right)^2$

FOR $0 < \eta < m^2/L$, $\gamma > 0$

TOTAL ITERATION COMPLEXITY: (DAMPED NEWTON)

PHASE I: $\dfrac{f(x_0) - f^*}{\gamma}$ TO BE COMPLETED (SUBLINEAR)

PHASE II: $\log\log\left(\frac{1}{\varepsilon}\right)$ TO ACHIEVE $f - f^* \leq \varepsilon$. (QUADRATIC)

THUS:

$$\boxed{\frac{f(x_0) - f^*}{\gamma} + \log\log\left(\frac{1}{\varepsilon}\right)}$$

(MOST DETAILS ARE OMITTED)

- SELF-CONCORDANCE.

- SHORTCOMINGS OF ANALYSIS FOR NEWTON'S METHOD
  1). COMPLEXITY ESTIMATES INVOLVE 3 CONSTANTS THAT ARE NEVER KNOWN IN PRACTICE.
  2) WHILE ALGORITHM IN PRACTICE IS AFFINE INVARIANT, THE ANALYSIS DEPENDS ON THE COORDINATE SYSTEM USED: IF WE CHANGE COORDINATES, M, m, L ALSO CHANGE.

- WE SEEK AN ALTERNATIVE TO THE ASSUMPTIONS:

$$mI \preceq \nabla^2 f(x) \preceq M \cdot I \quad \& \quad \|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L \|x-y\|_2$$

- NESTEROV & NEMIROVSKI: SELF-CONCORDANCE
  BARRIER FUNCTIONS $\rightarrow$ IPMS.
  i) BARRIER FUNCTIONS $\rightarrow$ IPMS.
  ii). NO UNKNOWN CONSTANTS IN ANALYSIS
  iii) AFFINE INVARIANCE.

ALTERNATIVE DEFINITION

$$\nabla^3 f(x)[h,h,h] \leq 2 \cdot \|h\|_x^3$$

WHERE

$$\|h\|_x = \sqrt{h^T \nabla^2 f(x) h}$$

- SELF-CONCORDANT FUNCTIONS ON $\mathbb{R}$

$$|f'''(x)| \leq 2 f''(x)^{3/2}$$

$$|f'''(x)| \leq k f''(x)^{3/2}$$

CAN BE TRANSFORMED TO

LINEAR + QUADRATIC FUNCTIONS. (CONVEX)

EXAMPLE 9.3.

KEY PROPERTY: SELF-CONCORDANCE IS AFFINE INVARIANT

- SELF-CONCORDANCE FOR FUNCTIONS ON $\mathbb{R}^n$

  $f: \mathbb{R}^n \to \mathbb{R}$ IS SELF-CONCORDANT IF IT IS SELF-CONCORDANT ALONG EVERY LINE: $f(x+tv)$, $\forall t$ & $\forall v \in \mathbb{R}^n$

- SELF-CONCORDANT CALCULUS
  - $|f_1''' + f_2'''| \leq \dots \leq 2(f_1''(x) + f_2''(x))^{3/2}$
  - COMPOSITION: E.G. f SELF-CONCORDANT $\rightarrow$ $f(Ax+b)$ IS ALSO.

  EXAMPLES: 9.4 - 9.6.

- PROPERTIES OF SELF-CONCORDANT FUNCTIONS.
  - UPPER AND LOWER BOUNDS ON SECOND DERIVATIVES. (UNDER CONVEXITY)

$$\frac{f''(0)}{(1+t\cdot f''(0)^{1/2})} \leq f''(t) \leq \frac{f''(0)}{(1-t\cdot f''(0))^{1/2}}$$

FOR $\forall t$ & $0 \leq t < f''(0)^{-1/2}$.

- THIS LEADS TO DIFFERENT LOWER & UPPER BOUNDS (BY INTEGRATING)

E.G., $\tilde{f}(t) = f(x+tv)$

AND

$$\tilde{f}(t) \geq \tilde{f}(0) + t\tilde{f}'(0) + t\tilde{f}''(0)^{1/2} - \log(1 + t\tilde{f}''(0)^{1/2})$$

SUCH CONDITIONS CAN BE USED IN ANALYSING NEWTON'S METHOD:

$$\frac{f(x_0) - f^*}{\gamma} + \log\log(1/\varepsilon) \longrightarrow$$ SIMILAR TO PREVIOUS ANALYSIS

WHERE γ DOES NOT DEPEND ON "WIERD" CONSTANTS.

$\boxed{0 \leq \gamma \leq 1/4}$ VS. $\boxed{0 \leq \gamma \leq m^2/L}$ (PREVIOUSLY).

INTERIOR POINT METHODS (IPMS)

ALSO, KNOWN AS BARRIER METHODS

- REASONS TO WORRY ABOUT IPMS.
  - SOLVE GENERAL CONVEX FORMULATIONS. (WITH CONSTRAINTS)
  - CONSTITUTE THE WORKHORSE OF TCS NEW RESULTS OR THE COMPARISON TO.

- CONSIDER THE FOLLOWING GENERAL FORMULATION:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \dots, m. \end{aligned} \qquad (*)$$

ALL FUNCTIONS ARE CONVEX AND TWICE DIFFERENTIABLE.
WE ASSUME THAT x^* EXISTS (PROBLEM IS SOLVABLE).

BY KKT CONDITIONS (+ SLATER'S CONSTRAINT QUALIFICATION)

$$\begin{aligned} g_i(x^*) &\leq 0, \quad i = 1, \dots, m \\ \mu_i^* &\geq 0, \quad i = 1, \dots, m \qquad (**) \\ \nabla f(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla g_i(x^*) &= 0. \\ \mu_i^* g_i(x^*) &= 0, \quad i = 1, \dots, m \end{aligned}$$

IPMS SOLVE $(*)$ (OR $(**)$) BY APPLYING NEWTON'S METHOD TO A SEQUENCE OF EQUALITY CONSTRAINED PROBLEMS, OR TO A SEQUENCE OF MODIFIED KKT CONDITIONS.

- THIS COURSE: PRIMAL, BARRIER-BASED, PATH FOLLOWING IPM.

  NOT IN THIS COURSE: PRIMAL-DUAL, VARIATIONS OF PRIMAL METHODS.

· LOGARITHMIC BARRIER FUNCTION AND CENTRAL PATH.

OUR GOAL IS TO FORMULATE THE PROBLEM SO THAT WE HANDLE THE INEQUALITY CONSTRAINTS. ONE WAY IS:

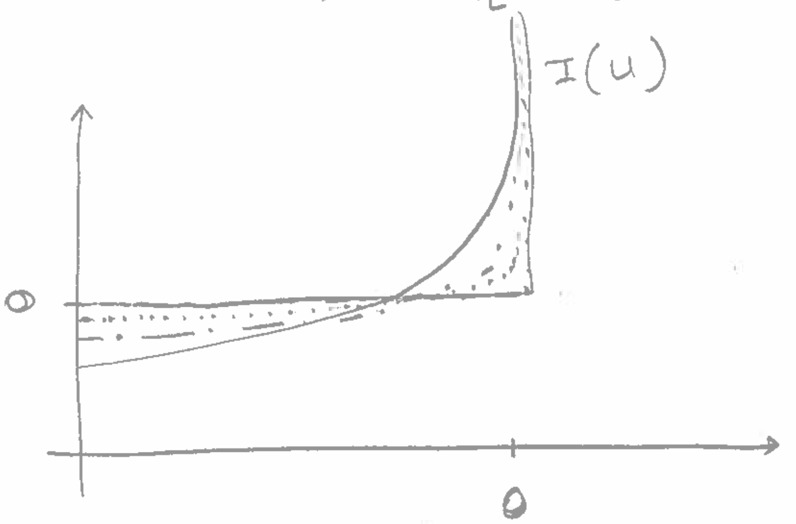
$$\min \; f(x) + \sum_{i=1}^{m} I_-\left(g_i(x)\right)$$

WHERE

$$I_-(u) = \begin{cases} 0, & u \leq 0 \\ \infty, & u > 0 \end{cases}$$

SINCE $I_-(\cdot)$ IS DISCONTINUOUS & NON-DIFFERENTIABLE, WE APPROXIMATE $I_-(\cdot)$ WITH:

$$\hat{I}_-(u) = -\frac{1}{t}\log(-u), \quad t > 0 \longrightarrow$$ SETS THE ACCURACY OF APPROXIMATION.



USING $\hat{I}_-(u)$:

$$\min \; f(x) + \underbrace{\sum_{i=1}^{m} \left(-\frac{1}{t}\right) \cdot \log(-g_i(x))}_{\text{CONVEX, FOR CONVEX } g_i(x)}$$

SOME DEFINITIONS:

$$\varphi(x) := -\sum_{i=1}^{m} \log(-g_i(x))$$: LOGARITHMIC BARRIER.

MAIN IDEA: CVX, pp. 564, TOP TWO-THREE PARAGRAPHS.

CENTRAL PATH

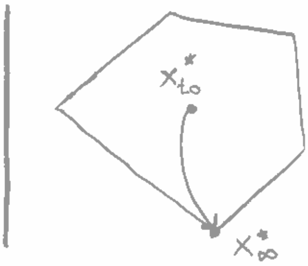
$$\min \; t \cdot f(x) + \varphi(x) \qquad (*)$$

THERE IS ALSO AN ALTERNATIVE DEFINITION: $\min f(x) + \varepsilon \cdot \varphi(x)$, $\varepsilon \to 0$

ASSUME FOR NOW THAT IS SOLVED VIA NEWTON'S METHOD, WITH A UNIQUE SOLUTION FOR EACH $t > 0$.

FOR $t > 0$, WE DEFINE $x^*(t)$ AS ITS SOLUTION.

CENTRAL PATH: SET OF POINTS $x^*(t)$, $\forall t > 0$.



POINTS ON CENTRAL PATH SATISFY DEFORMED KKT CONDITIONS:

$$g_i(x) \le 0 \quad i=1,\dots,m.$$

$$\mu_i \ge 0 \quad i=1,\dots,m$$

$$\nabla f(x) + \sum_{i=1}^{m} \mu_i \nabla g_i(x) = 0$$

$$\mu_i \, g_i(x) = -1/t \quad i=1,\dots,m$$

MOREOVER, USING DUALITY THEORY ONE CAN SHOW:

$$f(x^*(t)) - f^* \leq \frac{m}{t}, \quad m: \#\text{ CONSTRAINTS.}$$

I.E., $x^*(t)$ IS NO MORE THAN $\frac{m}{t}$ - SUBOPTIMAL.

- BARRIER METHOD.

SO FAR, WE ASSUMED WE CAN FIND $x^*(t)$. IN PRACTICE, IF WE WANT ACCURACY ε, WE CAN SET:

$$\varepsilon = m/t \Rightarrow t = m/\varepsilon. \quad \text{AND SOLVE:}$$

$$\min \left(\frac{m}{\varepsilon}\right) f(x) + \varphi(x) \quad \text{USING NEWTON'S METHOD.}$$

ALGORITHM: BARRIER METHOD (OR PATH-FOLLOWING METHOD)

INPUT: FEASIBLE x, $t := t_0 > 0$, $\mu > 1$, $\varepsilon > 0$.

1. COMPUTE $x^*(t)$ FROM
   $$\min\ t f(x) + \varphi(x) \longrightarrow \text{USUALLY, VIA NEWTON'S METHOD}$$
   STARTING FROM x.
2. $x := x^*(t)$
3. STOP IF $m/t < \varepsilon$.
4. $t := \mu t$

COMMENTS:

- ACCURACY OF STEP 1 COULD BE RELAXED: AFTER ALL, WE DO NOT SEARCH FOR $x^*(t)$, UNLESS $t \to \infty$. INEXACT $x^*(t)$ WILL STILL RESULT INTO A SEQUENCE OF SOLUTIONS TOWARDS $x^*(t)$ FOR $t \to \infty$. HOW MUCH INEXACT WE ARE AFFECTS COMPLEXITY
- CHOICE OF μ: SMALL $\mu \longrightarrow$ INITIAL POINT FROM PREVIOUS ITERATION COULD BE VERY CLOSE TO THE NEXT OPTIMAL POINT. (FEW NEWTON STEPS). BUT THIS LEADS TO MORE OUTER ITERATIONS (TRADE-OFF).
- CHOICE OF INITIAL t_0: t_0 LARGE $\longrightarrow$ FIRST OUTER ITERATION REQUIRES TOO MANY ITERATIONS

  t_0 SMALL $\longrightarrow$ EXTRA OUT ITERATIONS.

SHOW FIGURE 11.4 $\longrightarrow$ CVX, pp. 572

- CONVERGENCE ANALYSIS:

ACCURACY ε $\longrightarrow$ $\left\lceil \dfrac{\log \frac{m}{\varepsilon \cdot t_0}}{\log \mu} \right\rceil$ OUTER STEPS.

(ASSUMING NEWTON STEPS ARE FAST ENOUGH; $\log\log(1/\varepsilon)$)

(ARGUMENT: $\log\log(1/\varepsilon) \le 6$)

PP. 577 IN CVX BOOK.

QUESTION: AS t INCREASES, DOES NEWTON METHOD FOR CENTERING BECOME MORE DIFFICULT? TO BE ANSWERED..

- OVERALL ALGORITHM (TWO-PHASE ALGORITHM):

PHASE I: FIND $x^*(t_0)$ FOR ~~t_0~~ TO SOME ACCURACY.

PHASE II: APPLY BARRIER METHOD FOR INCREASING t.

- COMPLEXITY ANALYSIS VIA SELF-CONCORDANCE

PP. 585 IN CVX

ASSUMPTIONS:

- $tf + \varphi$ IS SELF-CONCORDANT $\forall t \ge t_0$.

NEWTON'S METHOD FOR SELF-CONCORDANT FUNCTIONS (INNER STEPS)

$$\frac{f(x) - f^*}{\gamma} + c \qquad \text{(IN GENERAL)}$$

WE START WITH t, $x^*(t)$

WE LOOK FOR μt, $x^*(\mu t)$

$$\frac{\mu t \cdot f(x) + \varphi(x) - \mu t \cdot f(x^*(\mu t)) - \varphi(x^*(\mu t))}{\gamma} + c$$

IS AN UPPER BOUND ON THE # OF ITERATIONS FOR NEWTON METHOD

UNFORTUNATELY, WE DO NOT KNOW $x^*(\mu t) \longrightarrow$ DERIVE ANOTHER UPPER BOUND

$$\mu t f(x) + \varphi(x) - \mu t \cdot f(x^*(\mu t)) - \varphi(x^*(\mu t)) \ldots$$
$$\le m(\mu - 1 - \log \mu)$$

USING DEFINITION OF $\varphi(x)$, SELF-CONCORDANCE, $\log \alpha \le \alpha - 1$ FOR $\alpha > 0$ AND DUALITY GAP.

THE CONCLUSION IS THAT:

$$\frac{m(\mu - 1 - \log\mu)}{\gamma} + c$$ IS AN UPPER BOUND ON # INNER ITERATIONS

DEPENDS ON: m, μ, $\gamma \to$ LINE SEARCH PARAMETER

$$\frac{1}{\gamma} = \frac{20 - 8\alpha}{\alpha\beta(1-2\alpha)^2}$$

DOES NOT DEPEND ON n.
— // — ON t.

TOTAL # OF NEWTON STEPS:

$$N = \left\lceil \frac{\log\left(\frac{m}{t_0 \varepsilon}\right)}{\log\mu} \right\rceil \cdot \left(\frac{m(\mu - 1 - \log\mu)}{\gamma} + c \right)$$ (AT LEAST LINEAR CONVERGENCE)

(SIMILAR ANALYSIS FOR PHASE I)

PP. 590 - CVX BOOK

REMARKS:

- CONSERVATIVE ANALYSIS VS. PRACTICE.
- ASSUMING SELF-CONCORDANCE, WE HAVE UNIFORM BOUND
- THOUGH, SELF-CONCORDANCE IS NOT NECESSARY.

- LONG-STEP VS SHORT-STEP IPMS

  - HOW WE UPDATE t PLAYS AN IMPORTANT ROLE IN COMPLEXITY.
    LARGE CHANGES $\longrightarrow$ MORE ITERATIONS IN INNER LOOP.
    SMALL — // — $\longrightarrow$ MORE — // — IN OUTER LOOP.
  - WHAT WE HAVE DESCRIBED SO FAR IS THE LONG-STEP APPROACH. WHERE # OF ITERATIONS IN INNER LOOP IS NOT EXACTLY SPECIFIED AND OFTEN UPPER BOUNDED.
  - CAN WE JUST PERFORM A SINGLE NEWTON STEP IN THE INNER LOOP? SHORT-STEP IPM!

DEFINITION: NEWTON'S DECREMENT:

$$\lambda_f(x) = \|\nabla f(x)\|_x^* \quad \text{FOR } f \text{ SELF-CONCORDANT AND}$$

$$\|g\|_x^* = \sup_{y:\|y\|_x} \langle g, y \rangle$$

(LONG STORY, SHORT)

- $\lambda_f(x)$ CAN BE USED TO IDENTIFY WHEN ONE IS IN THE QUADRATIC CONVERGENCE OF NEWTON'S METHOD FOR SELF-CONCORDANT FUNCTIONS.

THM. (NO PROOF).

3 LET f BE SELF-CONCORDANT; LET $x \in \text{INT}(\mathcal{X})$

LET. $\min_{x \in \mathcal{X}} f(x)$. THEN:

IF $\lambda_f(x) \leq 1/4$, THEN:

$$\lambda_f\left(x - \nabla^2 f(x)^{-1} \nabla f(x)\right) \leq 2 \lambda_f(x)^2$$

IN WORDS:

- IF INITIALIZED AT x_0 S.T. $\lambda_f(x_0) \leq 1/4$

  THEN: $\lambda_f(x_{k+1}) \leq 2 \cdot \lambda_f(x_k)^2$

4
- $\lambda_f(x_k) \leq 1/4$ PRESERVED.
- SOMEHOW DEFINES THE REGION OF QUADRATIC CONVERGENCE FOR SELF-CONCORDANT FUNCTIONS:

$$\{x : \lambda_f(x) \leq 1/4\}.$$

- DEFINITION OF ν-SELF-CONCORDANT BARRIERS:

5 $$\|\nabla \varphi(x)\|_x^* \leq \sqrt{\nu} \longrightarrow \langle \nabla \varphi(x), \nabla^2 \varphi(x)^{-1} \nabla \varphi(x) \rangle \leq \nu$$

$$|\nabla \varphi(x)^T h| \leq \nu^{1/2} \cdot \|h\|_x$$

- THEN, ONE-STEP IPMS:

$$t_{k+1} = \left(1 + \frac{1}{13\sqrt{\nu}}\right) t_k$$

FULL OBJECTIVE $tf + \varphi$

3 $$x_{k+1} = x_k - \nabla^2 F(x_k)^{-1} \nabla F(x_k)$$

WITH GUARANTEES: $O\left(\sqrt{\nu} \cdot \log \frac{\nu}{t_0 \varepsilon}\right)$ (NO WORSE THAN LONG-STEP)

(ALSO, ONE HAS TO WORRY ABOUT PHASE I/II ...)

- IN PRACTICE: TOO PESSIMISTIC; LONG-STEP PREFERRED