LECTURE 3

- UNCONSTRAINED VS CONSTRAINED OPT.

$$\min f(x) \quad \text{VS.} \quad \begin{array}{l}\min \; f(x) \\ \text{s.t.} \; x \in \mathcal{X}\end{array}$$

WE HAVE SEEN IN THE PAST SIMPLE CONSTRAINTS:

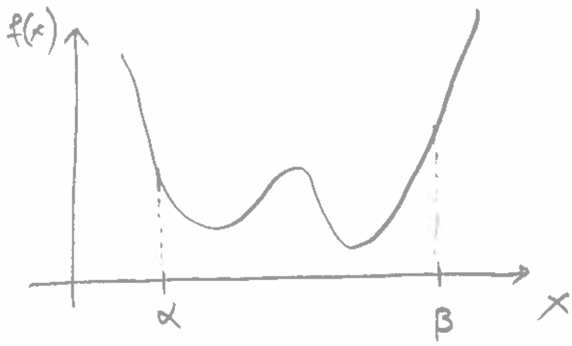
$$\left.\begin{array}{l} x \in \mathcal{X} \longrightarrow \|x\|_1 \leq \lambda \\ x \in \mathcal{X} \longrightarrow \|x\|_0 \leq k \\ x \in \mathcal{X} \longrightarrow \operatorname{rank}(X) \leq r \end{array}\right] \longrightarrow$$

SIMPLE CONSTRAINTS WITH SIMPLE PROJECTION OPERATIONS:

$$\Pi_C(y) = \underset{x \in \mathcal{X}}{\operatorname{argmin}} \; \frac{1}{2}\|x-y\|_2^2$$

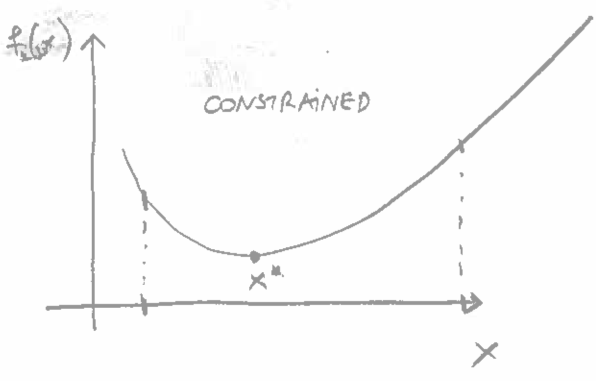
EXAMPLE:

$$\min \; f(x)$$
$$\text{s.t.} \; x \in [\alpha, \beta]$$

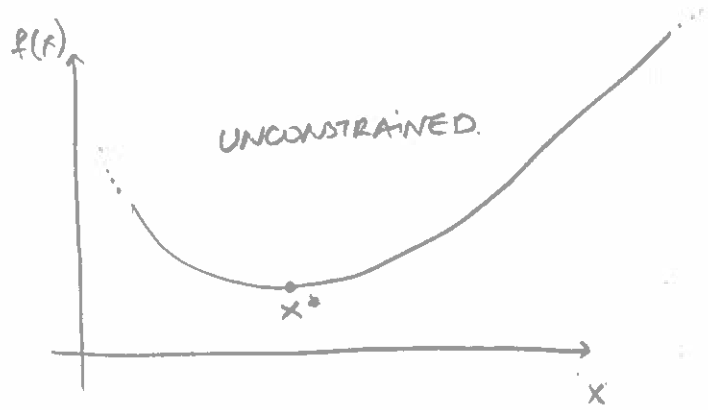


"COULD CONSTRAINTS BE "USELESS"? "

EXAMPLE:



VS.



GENERAL FORM OF CONSTRAINED PROBLEMS:

$$\min_{x \in \mathbb{R}^n} \; f(x)$$
$$\text{s.t.} \quad g_i(x) \leq 0, \quad i = 1, \ldots, m$$
$$\qquad h_i(x) = 0, \quad i = 1, \ldots, p.$$

- LAGRANGE MULTIPLIERS: SPECIAL CASE OF EQUALITY-CONSTRAINED OPT.

$$\min \; f(x)$$
$$\text{s.t.} \; h(x) = 0$$

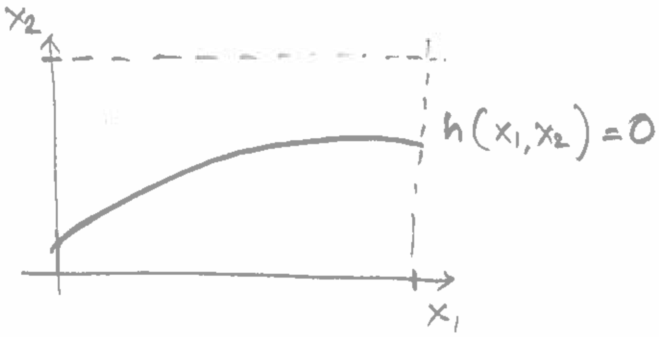
ASSUME THAT f, h HAVE CONTINUOUS PARTIAL DERIVATIVES

DUE TO THE CONSTRAINTS, CERTAINLY WE ARE INTERESTED ONLY IN THE POINTS IN THE FEASIBILITY SET:

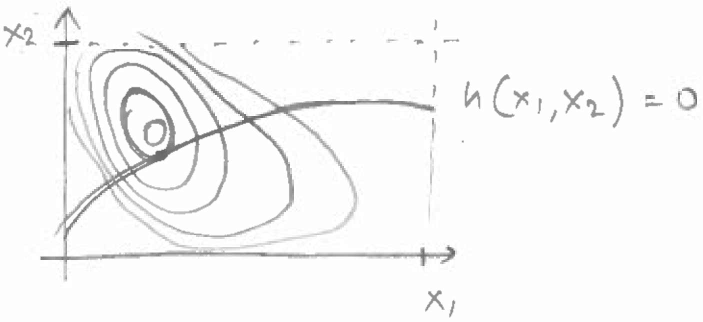
$$\{ x : h(x) = 0 \}$$

EXAMPLE:

2D-PROBLEM



LET US DRAW ON TOP OF THIS FIGURE THE CONTOUR LINES OF $f(\cdot)$



THE POINT WHERE $h(x_1, x_2) = 0$ TOUCHES THE MINIMUM CONTOUR IS THE POINT WHERE $f(\cdot)$ IS MINIMIZED S.T. $h(x_1, x_2) = 0$.

INTUITION: WE ARE INTERESTED IN FINDING POINTS WHERE f DOES NOT CHANGE, AS WE WALK ON $h(\cdot)$.
SUPPOSE WE ARE AT ANY POINT ON $h(x_1, x_2) = 0$, AS LONG AS WE MOVE AND f CHANGES, WE CAN KEEP MOVING AND FIND A NEIGHBORHOOD WHERE $f(\cdot)$ IS MINIMIZED.

THERE ARE TWO WAYS THIS COULD HAPPEN:

- WE COULD FOLLOW A CONTOUR OF f, AS WE MOVE ON $h(x_1, x_2) = 0$. (FIGURE 1 IN SLIDES)
- WE COULD FOLLOW A REGION OF f, WHERE f DOES NOT CHANGE AT ANY DIRECTION.

FOR THE FIRST CASE: CONTOUR LINES OF f AND h ARE PARALLEL:

$$\nabla f(\cdot) = \lambda \cdot \nabla h(\cdot)$$

(λ CONSTANT AS GRADIENTS MIGHT NOT BE OF ANY LENGTH.)

FOR THE SECOND CASE: IF f DOES NOT CHANGE IN ANY DIRECTION, THEN:

$$\nabla f(\cdot) = 0 \xrightarrow{\lambda = 0} \nabla f = \lambda \cdot \nabla h \text{ is SATISFIED.}$$

λ: LAGRANGE MULTIPLIER

LAGRANGIAN FUNCTION (BY REVERSE ENGINEERING THE ABOVE ARGUMENTS)

$$\mathcal{L}(x, \lambda) = f(x) - \lambda \cdot h(x) \quad \text{(ASK FOR INTUITION)}$$

SOLVING: $\nabla \mathcal{L}(x, \lambda) = 0 \longrightarrow$ WE SOLVE $\begin{cases} \nabla f = \lambda \cdot \nabla h \\ h(\cdot) = 0 \end{cases}$

SPECIFICALLY:

$$\nabla_x \mathcal{L}(x, \lambda) = 0 \Rightarrow \nabla f(x) - \lambda \nabla h(x) = 0 \Rightarrow \nabla f(x) = \lambda \cdot \nabla h(x)$$

$$\nabla_\lambda \mathcal{L}(x, \lambda) = 0 \Rightarrow 0 - h(x) = 0 \Rightarrow h(x) = 0$$

STATIONARY POINT: COULD BE LOCAL MINIMA
—"— MAXIMA
SADDLE POINT

IF WE MAKE ASSUMPTIONS ON f, h, THEN WE COULD GUARANTEE GLOBAL OPTIMALITY.

(NOTE: THERE IS NOTHING ALGORITHMIC HERE SO FAR)

EXAMPLE: $f(x_1, x_2) = -\exp\left(-\left(x_1 x_2 - \frac{3}{2}\right)^2 - \left(x_2 - \frac{3}{2}\right)^2\right)$

$h(x_1, x_2) = 0 \Rightarrow x_1 - x_2^2 = 0.$

DEFINE: $\mathcal{L}(x_1, x_2, \lambda) = f(x_1, x_2) - \lambda \cdot h(x_1, x_2)$

COMPUTE: $\frac{\partial \mathcal{L}}{\partial x_1} = 2 \cdot x_2 \cdot f(x_1, x_2)\left(\frac{3}{2} - x_1 x_2\right) - \lambda$

$$\frac{\partial \mathcal{L}}{\partial x_2} = f(x)\left(-2x_1\left(x_1 x_2 - \frac{3}{2}\right) - 2\left(x_2 - \frac{3}{2}\right)\right) + 2\lambda x_2$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = x_2^2 - x_1$$

(FIGURE 2 IN SLIDES)

SETTING TO ZERO: $x_1 \approx 1.358$, $x_2 \approx 1.165$, $\lambda \approx 0.17$

MULTIPLE EQUALITY CONSTRAINTS

$$\min \; f(x)$$
$$\text{s.t.} \quad h_1(x) = 0$$
$$h_2(x) = 0$$
$$\vdots$$
$$h_p(x) = 0$$

IS THERE A PARTICULAR REASON WE ASSUME "-"?

THEN, THE LAGRANGIAN TAKES THE FORM:

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i=1}^{p} \lambda_i \cdot h_i(x) = f(x) - \lambda^T \cdot \vec{h}(x)$$

(EXAMPLE ON SLIDES)

WHERE $\vec{h}(x)$ ENCAPSULATES ALL $h_i(x)$ AS A VECTOR

- LAGRANGE MULTIPLIERS: SPECIAL CASE OF INEQUALITY CONSTRAINTS

$$\min \; f(x)$$
$$\text{s.t.} \quad g(x) \leq 0$$

INCONVENIENT FORM: $\min \; f_\infty(x) := \begin{cases} f(x), & \text{if } g(x) \leq 0 \\ \infty, & \text{otherwise} \end{cases}$

$$= f(x) + \infty(g(x) > 0)$$

LAGRANGIAN FORM: $\mathcal{L}(x, \mu) = f(x) + \mu \cdot g(x)$ (FIGURE 3 IN SLIDES)

WHICH APPROXIMATES $f_\infty(x)$ AS IN.

$$f_\infty(x) = \max_{\mu \geq 0} \mathcal{L}(x, \mu)$$

COMBINING THE ABOVE: $\min_x \left(\max_{\mu \geq 0} \mathcal{L}(x, \mu) \right)$

- GENERAL LAGRANGIAN FUNCTION

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \underbrace{\sum_{i=1}^{m} \mu_i g_i(x)}_{\text{INEQUALITY CONS.}} + \underbrace{\sum_{i=1}^{p} \lambda_i \cdot h_i(x)}_{\text{EQUALITY CONSTR.}}$$

λ, μ: DUAL VARIABLES OR LAGRANGE MULTIPIER VECTORS.

FOCUSING ON (λ, μ), WE CAN DEFINE THE DUAL FUNCTION:

$$q(\lambda,\mu) = \inf_x \mathcal{L}(x,\lambda,\mu) = \inf_x \left(f(x) + \sum_{i=1}^{m} \mu_i \cdot g_i(x) + \sum_{i=1}^{p} \lambda_i \cdot h_i(x) \right)$$

WHAT KIND OF FUNCTION (W.R.T. (λ,μ)) IS $q(\lambda,\mu)$? CONCAVE

DOES IT DEPEND ON g, h, f? NO

POINTWISE INFIMUM OF AFFINE FUNCTIONS:

(EXAMPLE ON SLIDES)

LOWER BOUNDS ON OPTIMAL VALUE

LET f^* BE THE MINIMUM OBJ. VALUE OF

$$\begin{aligned} \min\ & f(x) \\ \text{s.t. } & g(x) \le 0 \\ & h(x) = 0. \end{aligned} \qquad (*)$$

THEN, FOR ANY $\mu \geq 0$ AND λ:

$$q(\lambda,\mu) \le f^*$$ (FIGURE 4 ON SLIDES)

PROOF: LET $\tilde{x}$ BE A FEASIBLE POINT FOR $(*)$, I.E., $g_i(\tilde{x}) \le 0$

$h_i(\tilde{x}) = 0$

$\mu \geq 0$.

THEN: $\sum_{i=1}^{m} \mu_i \cdot g_i(\tilde{x}) + \sum_{i=1}^{p} \lambda_i \cdot h_i(\tilde{x}) \le 0.$

THEREFORE: $\mathcal{L}(\tilde{x},\lambda,\mu) = f(\tilde{x}) + \sum_{i=1}^{m} \mu_i \cdot g_i(\tilde{x}) + \sum_{i=1}^{p} \lambda_i \cdot h_i(\tilde{x})$

$\le f(\tilde{x})$

HENCE: $q(\lambda,\mu) = \inf_x \mathcal{L}(x,\lambda,\mu) \le \mathcal{L}(\tilde{x},\lambda,\mu) \le f(\tilde{x}), \ \forall \tilde{x}$ ∎

CONNECTION TO INDICATOR FUNCTIONS:

$$\min_x \ f(x) + \sum_{i=1}^{m} I_-(g_i(x)) + \sum_{i=1}^{p} I_0(h_i(x))$$

WHERE: $I_-(u) = \begin{cases} 0, & u \le 0 \\ \infty, & u > 0. \end{cases}$

SIMILARLY $I_0(\cdot)$.

WITH THE DUAL FUNCTION, WE ~~APPROXIMATE~~ THE INDICATOR FUNCTION WITH A LINEAR FUNCTION

EXAMPLES: pp. 218 - 221

LS SOLUTION OF LINEAR EQUATIONS

STANDARD FORM LP $\longrightarrow$ CVX BOOK

- THE LAGRANGE DUAL PROBLEM.

$\forall (\lambda, \mu)$ WITH $\mu \geq 0$: LAGRANGE DUAL FUNCTION GIVES A LOWER BOUND ON THE OPTIMAL f^*

BEST LOWER BOUND?

$$\max \; q(\lambda, \mu) \quad \longrightarrow \quad " \quad "$$
$$\text{s.t.} \; \mu \geq 0$$

EXAMPLE: pp. 224: LP

pp. 225: INEQUALITY LP.

- WEAK DUALITY: $d^* \leq f^*$

$f^* - d^*$: OPTIMAL DUALITY GAP

- STRONG DUALITY: $d^* = f^*$

- STRONG DUALITY DOES NOT GENERALLY HOLD, UNLESS PRIMAL IS CONVEX.

EVEN IF PRIMAL IS CONVEX $\not\rightarrow$ STRONG DUALITY.

- MIN-MAX CHARACTERIZATION (ONLY INEQUALITY FOR CLARITY)

$$f^* = \inf_{x} \sup_{\mu \geq 0} \mathcal{L}(x, \mu)$$

$$d^* = \sup_{\mu \geq 0} \inf_{x} \mathcal{L}(x, \mu)$$

WEAK DUALITY: $\sup_{\mu \geq 0} \inf_{x} \mathcal{L}(x, \mu) \leq \inf_{x} \sup_{\mu \geq 0} \mathcal{L}(x, \mu)$

$\downarrow$

STRONG DUALITY $\equiv$
(WE CAN SWITCH MIN-MAX)

ALSO, SADDLE POINT INTERPRETATION $\longrightarrow$ GANS, GAME THEORY

- OPTIMALITY CONDITIONS

— COMPLEMENTARY SLACKNESS

ASSUME STRONG DUALITY: $d^* = f^*$. THEN:

$$\begin{aligned} f^* &= q(\lambda^*, \mu^*) \\ &= \inf_x \left(f + \sum_{i=1}^m \mu_i^* \cdot g_i(x) + \sum_{i=1}^p \lambda_i^* \cdot h_i(x) \right) \\ &\le f(x^*) + \sum_{i=1}^m \mu_i^* g_i(x^*) + \sum_{i=1}^p \lambda_i^* \cdot h_i(x^*) \\ &\overset{\mu_i^* \ge 0}{\underset{g_i(x^*) \le 0}{\le}} f(x^*) = f^* \end{aligned}$$

CONCLUSION: $\sum_{i=1}^m \mu_i^* g_i(x^*) = 0 \Rightarrow \mu_i^* \cdot g_i(x^*) = 0, \ \forall i$

(COMPLEMENTARY SLACKNESS)

$$\mu_i^* > 0 \Rightarrow g_i(x^*) = 0 \quad \text{OR} \quad g_i(x^*) < 0 \Rightarrow \mu_i^* = 0$$

- KKT OPTIMALITY CONDITIONS. (EVEN FOR NONCONVEX PROBLEMS)

LET x^* AND (λ^*, μ^*) BE ANY PRIMAL AND DUAL OPTIMAL POINTS WITH ZERO DUALITY GAP. SINCE x^* MINIMIZES $\mathcal{L}(x, \lambda^*, \mu^*)$ OVER X:

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

THUS:

$$\left.\begin{aligned} g_i(x^*) &\le 0, & i &= 1, \dots, m \\ h_i(x^*) &= 0, & i &= 1, \dots, p \\ \mu_i^* &\ge 0, & i &= 1, \dots, m. \\ \mu_i^* g_i(x^*) &= 0, & i &= 1, \dots m \\ \nabla f(x^*) + \sum \mu_i^* \nabla g_i(x^*) &+ \sum \lambda_i h_i(x^*) = 0 \end{aligned}\right]$$

KARUSH-~~KUHN~~-TUCKER CONDITIONS

IN WORDS, ANY OPT. PROBLEM WITH DIFFERENTIABLE OBJECTIVE + CONTRAINT FUNCTIONS FOR WHICH STRONG DUALITY HOLDS, ANY PAIR OF PRIMAL/DUAL OPTIMAL POINTS SATISFY KKT CONDITIONS.

- KKT OPTIMALITY CONDITIONS (CONVEX PROBLEMS)

UNDER CONVEXITY: KKT CONDITIONS ARE ALSO SUFFICIENT FOR POINTS TO BE OPTIMAL (PRIMAL/DUAL). I.E., ANY POINTS $\tilde{x}, \tilde{\mu}, \tilde{\lambda}$ THAT SATISFY KKT, ARE PRIMAL/DUAL OPTIMAL, WITH ZERO DUALITY GAP.

PROOF: THE FIRST TWO CONDITIONS $\longrightarrow$ $\tilde{x}$ IS PRIMAL FEASIBLE.

$\tilde{\mu}_i \geq 0 \longrightarrow \mathcal{L}(x, \tilde{\mu}, \tilde{\lambda})$ IS CONVEX IN x

THE LAST CONDITION $\longrightarrow$ $\nabla_x \mathcal{L}(\cdot) = 0$ WHEN $x = \tilde{x}$

THUS, $\tilde{x}$ MINIMIZES $\mathcal{L}(x, \tilde{\mu}, \tilde{\lambda})$ OVER x.

WE CONCLUDE:

$$
\begin{aligned}
q(\tilde{\mu}, \tilde{\lambda}) &= \mathcal{L}(\tilde{x}, \tilde{\mu}, \tilde{\lambda}) \\
&= f(\tilde{x}) + \sum_{i=1}^{m} \tilde{\mu}_i \, g_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\lambda}_i \cdot h_i(\tilde{x}) \\
&= f(\tilde{x})
\end{aligned}
$$

WHERE WE USED: $h_i(\tilde{x}) = 0$ & COMPLEMENTARY SLACKNESS $\longrightarrow$ ZERO DUALITY GAP.

- KKT CONDITIONS TOWARDS FINDING SOLUTION

IT IS POSSIBLE TO SOLVE THE KKT CONDITIONS ANALYTICALLY, TO SOLVE THE ORIGINAL PROBLEM.

MANY ALGORITHMS IN CONVEX OPT. ARE CONCEIVED AS METHODS FOR SOLVING THE KKT CONDITIONS.

EXAMPLE 5.1, PP. 244

- SOLVING PRIMAL VIA DUAL

ASSUME STRONG DUALITY; ASSUME (μ^*, λ^*) IS KNOWN. SUPPOSE THAT THE MINIMIZER OF $\mathcal{L}(x, \mu^*, \lambda^*)$ IS UNIQUE (STRICTLY CONVEX):

$$
\min \; f(x) + \sum_{i=1}^{m} \mu_i^* g_i(x) + \sum_{i=1}^{p} \lambda_i^* h_i(x) \qquad (*)
$$

THEN SOLVING $(*)$ MEANS THAT WE SOLVE THE ORIGINAL CONSTRAINED PROBLEM, VIA THE UNCONSTRAINED $(*)$.

EXAMPLE 5.3, PP. 248.

FROM DUALITY, TO DUAL ASCENT, TO AUGMENTED LAGRANGIAN TO ADMM.

ASSUME: $\min_x f(x)$ s.t. $Ax = b$, $A \in \mathbb{R}^{m \times n}$ $\longrightarrow$ $\min_x f(x)$ s.t. $h(x) = Ax - b = 0$

$\downarrow$

$\min f(x)$ s.t. $h_i(x) = 0$ FOR $h_i(x) = \alpha_i^T x - b_i$

LET US FORM THE LAGRANGIAN:

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i \cdot h_i(x)$$

$$= f(x) + \sum_{i=1}^{m} \lambda_i \cdot (\alpha_i^T x - b_i)$$

$$= f(x) + \lambda^T (Ax - b)$$

LET US FORM THE DUAL FUNCTION:

$$q(\lambda) = \inf_x \mathcal{L}(x, \lambda)$$

WITH THE DUAL PROBLEM BEING

$$(*) \quad \sup_\lambda q(\lambda). \longleftarrow$$ WHAT PROBLEM IS THIS: CVX OR NON-CVX?

LET US TRY TO SOLVE $(*)$ WITH GRADIENT ASCENT:

1. SET UP INITIAL $\lambda_0 \in \mathbb{R}^m$

(INTUITIVELY, WE WOULD LIKE TO PERFORM: $\lambda_{t+1} = \lambda_t + \eta \cdot \nabla q(\lambda_t)$)

2. HOWEVER, $q(\cdot) = \inf_x \mathcal{L}(x, \cdot)$. ASSUME WE CAN SOLVE:

$$\inf_x \mathcal{L}(x, \lambda_t) \longrightarrow q(\lambda_t) \text{ FOR } x_t^*$$

3. COMPUTE $\nabla_{\lambda_t} q(\lambda_t) = \nabla_{\lambda_t} \left(f(x_t) + \lambda_t^T (Ax_t - b) \right)$

$$= 0 + (Ax_t - b)$$

4. UPDATE, $\lambda_{t+1} = \lambda_t + \eta \nabla q(\lambda_t) = \lambda_t + \eta (Ax_t - b)$ AND REPEAT.

EVEN IF $q(\lambda)$ IS NOT STRONGLY CONVEX, WE CAN MAKE IT STRONGLY CONVEX BY INCORPORATING A PROXIMAL TERM.

INSTEAD:

$$\sup_\lambda q(\lambda) = \sup_\lambda \inf_x \mathcal{L}(x,\lambda)$$
$$= \sup_\lambda \inf_x \left(f(x) + \lambda^T(Ax-b) \right)$$

WE DO:

WHY "−"

$$\sup_\lambda \inf_x \left(f(x) + \lambda^T(Ax-b) - \frac{1}{2\eta_t} \|\lambda - \lambda_t\|_2^2 \right)$$

UNDER STRONG DUALITY: $\sup \inf = \inf \sup$.

$$\inf_x \sup_\lambda \left(f(x) + \lambda^T(Ax-b) - \frac{1}{2\eta_t} \|\lambda_t - \lambda\|_2^2 \right)$$
$$= \inf_x \left(f(x) + \lambda_t^T (Ax-b) + \frac{\eta_t}{2} \|Ax-b\|_2^2 \right)$$

WHERE THE INNER SUP. IS OPTIMIZED IN CLOSED-FORM BY

$$\lambda = \lambda_t + \eta_t (Ax-b)$$

<u>DEFINITION</u>: AUGMENTED LAGRANGIAN IS:

DIFFERENTIABLE UNDER MILDER CONDITIONS

$$\mathcal{L}_\eta(x,\lambda) = f(x) + \lambda^T(Ax-b) + \frac{\eta}{2}\|Ax-b\|_2^2$$

CAN BE VIEWED AS THE LAGRANGIAN OF:

$$\min_x f(x) + \frac{\eta}{2}\|Ax-b\|_2^2 \quad \text{s.t. } Ax = b$$

THEN, THE <u>AUGMENTED</u> LAGRANGIAN METHOD IS:

$$x_t = \inf_x \mathcal{L}_{\eta_t}(x, \lambda_t)$$

$$\lambda_{t+1} = \lambda_t + \eta_t (Ax_t - b)$$

CONVERGES UNDER FAR MORE GENERAL CONDITIONS THAN DUAL ASCENT (NON-STRICT CONVEX, $f \to +\infty$)

WHILE SIMILAR TO DUAL ASCENT:

- AUGMENTED LAGRANGIAN CAN SPEED UP CONVERGENCE, BUT x_t INVOLVES $\frac{\eta_t}{2}\|Ax-b\|^2$ TERM, WHICH MIGHT BE DIFFICULT TO HANDLE.
- NEVERTHELESS, $O(1/t)$ RATE (IMPROVED OVER DUAL ASCENT)

DUAL DECOMPOSITION: DUALITY CAN LEAD TO TRIVIAL PARALLELIZATION.

CONSIDER:

$$x = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(N)} \end{bmatrix}, \quad x^{(i)} \in \mathbb{R}^{n_i} \quad \text{FOR} \quad \sum_{i=1}^{N} n_i = n$$

$$A = \left[A_1 \,|\, A_2 \,|\, \dots \,|\, A_N \right] \quad \text{SUCH THAT} \quad Ax = \sum_{i=1}^{N} A_i x^{(i)}$$

$$f(x) = \sum_{i=1}^{N} f_i\left(x^{(i)}\right)$$

THEN:

$$\mathcal{L}(x, \lambda) = \sum_{i=1}^{N} \left(f_i\left(x^{(i)}\right) + \lambda^T A_i x^{(i)} - \frac{1}{N} \lambda^T b \right)$$

$$= \sum_{i=1}^{N} \mathcal{L}_i\left(x^{(i)}, \lambda\right)$$

$\hookrightarrow$ NON-INTERACTING PARTITIONS $\left(x^{(i)}, A_i, f_i\right)$

ALGORITHM:

- IN PARALLEL (WORKERS): $x_t^{(i)} = \inf_{x^{(i)}} \mathcal{L}_i\left(x^{(i)}, \lambda_t\right)$
- SYNCHRONIZE (MASTER): $\lambda_{t+1} = \lambda_t + \eta(Ax - b)$.

EXAMPLES: CONSENSUS OPT., NETWORK UTILITY MAXIMIZATION, ..

HOWEVER, FOR THE DUAL DECOMPOSITION TO WORK, WE USE DUAL ASCENT, NOT AUGMENTED LAGRANGIAN. WHY? $\|Ax - b\|^2$ IS NOT STRAIGHTFORWARDLY DECOMPOSED IN x.

ADMM: ALTERNATING DIRECTION METHOD OF MULTIPLIERS

$\hookrightarrow$ PARALLESISM $\boxed{\checkmark}$

$\hookrightarrow$ FASTER CONVERGENCE $\boxed{\checkmark}$

CONSIDER THE PROBLEM:

$$\min_{x,z} \quad f(x) + g(z)$$
$$\text{s.t.} \quad Ax + Bz \leq c$$

$\longrightarrow$ WE SPLIT OBJECTIVE INTO + CONSTRAINTS

TWO BLOCKS, x & z.

WHILE AUGMENTED LAGRANGIAN WOULD SOLVE:

$$(x_{t+1}, z_{t+1}) = \inf_{x,z} \mathcal{L}_\eta(x, z, \lambda_t)$$
$$\lambda_{t+1} = \lambda_t + \eta\,(Ax_{t+1} + Bz_{t+1} - c)$$

IN ADMM, WE DO:

$$x_{t+1} = \inf_{x} \mathcal{L}_\eta(x, z_t, \lambda_t)$$
$$z_{t+1} = \inf_{z} \mathcal{L}_\eta(x_{t+1}, z, \lambda_t)$$

IS THIS PARALLELIZABLE?

NO: z_{t+1} WAITS FOR x_{t+1}....

$$\lambda_{t+1} = \lambda_t + \eta\,(Ax_{t+1} + Bz_{t+1} - c)$$

USUALLY, INSTEAD OF CONVERGENCE RATE, WE ONLY ACHIEVE AN ASYMPTOTIC CONVERGENCE GUARANTEE.

$$\left(\mathcal{L}_\eta(x, z, \lambda) = f(x) + g(z) + \lambda^T (Ax + Bx - c) + \frac{\eta}{2} \| Ax + Bz - c \|_2^2 \right)$$

<u>Q</u>: CAN WE USE ADMM WITH ADAM AS A SUBSOLVER?

<u>Q</u>: WHAT IF WE COMPUTE:

$$x_{t+1} = \inf_{x} \mathcal{L}_\eta(x, z_t, \lambda_t)$$
$$z_{t+1} = \inf_{z} \mathcal{L}_\eta(x_t, z, \lambda_t)$$

?

* READING ASSIGNMENTS:

i) USEFULNESS OF DUAL METHODS

ii) CONVERGENCE OF ADMM.

iii) EFFICIENT DUAL METHODS IN MACHINE LEARNING