$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t. } h(x) = 0$$

$\longrightarrow$ Augmented Lagrangian

$$L_c(x,\nu) = f(x) + \nu^T h(x) + \frac{c}{2}\|h(x)\|^2$$

$\nu \to \nu^*$ and $c \to \infty$ $\Rightarrow$ $x \to x^*$

Dual decomposition $\to$ ADMM

At iteration k: (Method of Multipliers)

$$\boxed{\begin{aligned} x_k &= \arg\min L_{c_k}(x, \nu_k) \to x^* \\ \nu_{k+1} &= \nu_k + c_k h(x_k) \to \nu^* \\ c_{k+1} &= \beta c_k, \ \beta > 1 \quad \beta \in (5,10) \end{aligned}}$$

$$h(x_k) = \frac{\nu_{k+1} - \nu_k}{c_k}$$

$$L_{c_k}(x,\nu_k) = f(x) + \nu_k^T h(x) + \frac{c_k}{2}\|h(x)\|^2$$

① Step 1: $\nabla_x L_{c_k}(x,\nu_k) \to 0$

$$\longrightarrow \nabla f(x) + \nabla h(x)\left(\nu_k + c_k h(x)\right) = 0$$

Suppose you consider the original Lagrangian:

$$L(x,\nu) = f(x) + \nu^T h(x)$$

$$\nabla_x L = 0 \Rightarrow \nabla f(x^*) + \nu^{*T} \nabla h(x^*) = 0$$

Proposition: Let f and h are continuous functions, and the constraint set $\{h(x)=0\}$ is non-empty.
For $k = 0,1,2,\dots$, let x_k be the global min of the following optimization problem,

$$\min_{x \in \mathbb{R}^n} L_{c_k}(x,\nu_k),$$

where $\{\nu_k\}$ is bounded, $0 < c_k < c_{k+1}$ $\forall k$ with $c_k \to \infty$.
Then every limit point of the seq. $\{x_k\}$ is a global minimum of the original problem.

* Caveat: Assumes that we can do exact minimization of the augmented Lagrangian.

* Proposition: Let x_k for $k=0,1,\dots,$ satisfies.

$$\|\nabla_x L_{c_k}(x_k, \nu_k)\| \leq \varepsilon_k \; ,$$

where $\{\nu_k\}$ is bounded, and let $\{\varepsilon_k\}$ and $\{c_k\}$ satisfy.

$$0 < c_k < c_{k+1}, \forall k \text{ with } c_k \to \infty, \quad 0 \leq \varepsilon_k \; \forall k \text{ with } \varepsilon_k \to 0,$$

then $x_k \to x^*$, where x^* is such that $\nabla h(x^*)$ has rank m, and then $\nu_k + c_k h(x_k) \to \nu^*$.

Algorithm is fairly robust, but there are certain practical issues.

Ex: $f(x) = \frac{1}{2}(x_1^2 + x_2^2)$

$$\min_{(x_1,x_2)} f(x) \quad \text{s.t.} \quad x_1 = 1$$

$(x_1^*, x_2^*) = (1,0)$

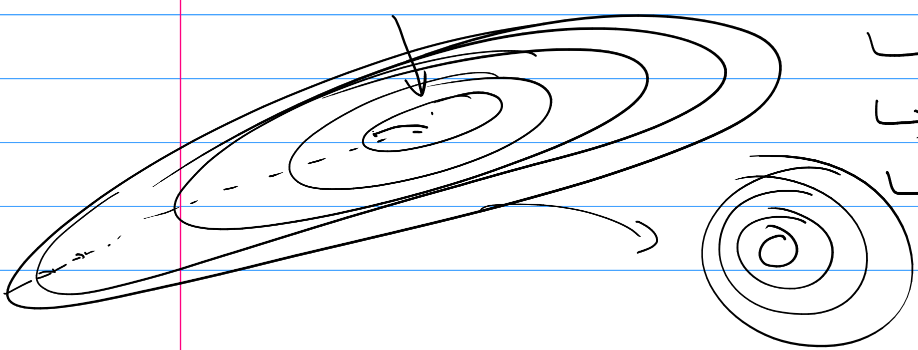
$\nu^* = -1$

$$L_c(x,\nu) = \frac{1}{2}(x_1^2+x_2^2) + \frac{c}{2}(x_1-1)^2 + \nu(x_1-1)$$

$$\nabla_x L_c(x,\nu) = \begin{bmatrix} x_1 + c(x_1-1) + \nu \\ x_2 \end{bmatrix}$$

$$\nabla^2_{xx} L_c(x,\nu) = \begin{bmatrix} 1+c & 0 \\ 0 & 1 \end{bmatrix}$$

$\to$ As $c \to \infty$ matrix is ill-conditioned



$\hookrightarrow$ Newton-type algorithm

$\hookrightarrow$ Use good starting points.

$\hookrightarrow$ Increase c_k at a moderate rate.

$$x_k = \underline{\arg\min} \; L_{c_k}(x, \nu_k)$$

$x_{2,k} = 0 \qquad x_{1,k} + \nu_k + c_k(x_{1,k} - 1) = 0$

$$\boxed{x_{1,k} = \frac{c_k - \nu_k}{1 + c_k}}$$

$$\nu_{k+1} = \nu_k + c_k h(x_k)$$

$$= \nu_k + c_k (x_{1,k} - 1)$$

$$= \nu_k + c_k \left(\frac{c_k - \nu_k}{1 + c_k} - 1 \right) = \nu_k + c_k \left(\frac{-1 - \nu_k}{1 + c_k} \right)$$

$$\nu_{k+1} = \nu_k - c_k \left(\frac{1 + \nu_k}{1 + c_k} \right)$$

$$\underline{\nu_{k+1} - \nu^*} = \frac{\nu_k - \nu^*}{\underline{1 + c_k}} \qquad c_k > 0$$

<u>Ex:</u> Non-convex problem.

$$\min_{x_1, x_2} \frac{1}{2}\left(\underline{-x_1^2} + x_2^2\right) \qquad (x_1^*, x_2^*) = (1, 0)$$

$$\text{s.t.} \quad x_1 = 1 \qquad \nu^* = 1$$

$$x_k = \arg\min_x L_{c_k}(x, \nu_k)$$

$$x_k = \left(\frac{c_k - \nu_k}{c_k - 1}, 0 \right) \xrightarrow[\text{as } c_k \to \infty]{} x^* = (1, 0)$$

$$\nu_{k+1} - \nu^* = \frac{\nu_k - \nu^*}{c_k - 1}$$, we must chose $c_k > 2$

$\bar{c}$

Algorithm is very robust.

↳ No need to increase $c_k \to \infty$ for convergence as long as $c_k > 2$.

$\ast$ Inequality constrained optimization problems:

$$\min_{x\in\mathbb{R}^n} f(x)$$

$$\text{s.t.} \quad h(x) = 0$$

$$\boxed{g_j(x) \le 0} \quad \forall j \in \{1, 2, \ldots, r\}$$

$$g_j(x) + z_j^2 = 0 \quad \forall j \in \{1, 2, \ldots, r\}$$

$$\boxed{\begin{aligned} L_c(x, z, \nu, \lambda) = {} & f(x) + \nu^T h(x) + \frac{c}{2}\|h(x)\|^2 \\ & + \frac{c}{2}\sum_{j=1}^{r} |g_j(x) + z_j^2|^2 \\ & + \sum_{j=1}^{r} \lambda_j \left(g_j(x) + z_j^2\right) \end{aligned}}$$

Method of Multipliers:

① $(x_k, z_k) \leftarrow \arg\min_{(x,z)} L_{c_k}(x, z, \nu_k, \lambda_k)$

② $\nu_{k+1} = \nu_k + c_k h(x_k) \rightarrow \nu^*$

③ $\lambda_{j,k+1} = \lambda_{j,k} + c_k\left(g_j(x_k) + z_{j,k}^2\right)$

④ $c_{k+1} = \beta c_k$

$$\max\{0, \lambda_{j,k} + c_k(g_j(x_k))\} \rightarrow \lambda_j^*$$

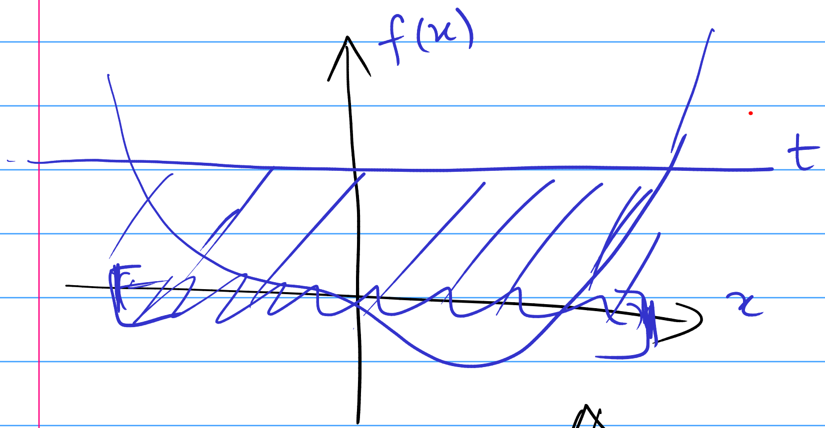
X ——— X ——— X

Dual ascent and dual decomposition

* Conjugate of a function f

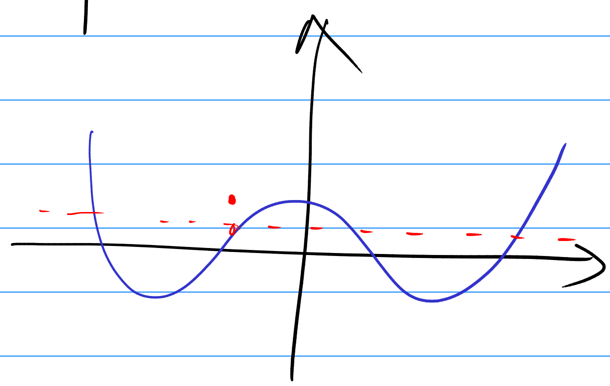
$$f^*(y) := \max_{x \in \mathbb{R}^n} \{ y^T x - f(x) \}$$

Thm: If f is closed and convex, then $f^{**} = f$.



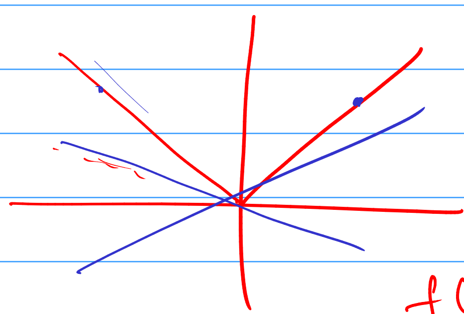
Epigraph of a function

$$\text{epi}(f) = \{ (x,t) \in \mathbb{R}^{n+1} \mid x \in \text{dom } f,\ f(x) \le t \}$$



$$x \in \partial f^*(y) \iff y \in \partial f(x) \iff x \in \arg\min_z \{ f(z) - y^T z \}$$

$x = \nabla f^*(y)$



$y = \nabla f(x)$, then $x = \nabla f^*(y)$

Subgradients

$$f(y) \ge f(x) + \nabla f(x)^T (y - x)$$

g

$g = \{ [-1, 1] \}$

Primal problem:

$$\min_x f(x)$$
$$\text{s.t.} \quad Ax = b$$

Dual Problem:

$$\max_\nu \underbrace{-\nu^T b - f^*(-A^T\nu)}_{g(\nu)}$$

$$\nabla g(\nu) = -b + A \underbrace{\nabla f^*(-A^T\nu)}_{x}$$

$$x \leftarrow \arg\min_z \{ f(z) + \nu^T A z \}$$

$$L(x,\nu) = f(x) + \nu^T(Ax - b)$$

$$g(\nu) = \min_x L(x,\nu) = -\nu^T b + \min_x \{ f(x) + \nu^T A x \}$$

At iteration k

$$\boxed{\begin{aligned} &(1)\quad x_k = \arg\min L(x, \nu_k) \\ &(2)\quad \nabla g(\nu_k) = A x_k - b \\ &(3)\quad \nu_{k+1} = \nu_k + \alpha_k \nabla g(\nu_k) \end{aligned}} \rightarrow \text{Dual ascent}$$

∗ Convergence guarantees:

↳ If f is μ-SC $\Rightarrow$ then with a constant step-size $\alpha_k = \mu$

↳ $O\left(\frac{1}{\varepsilon}\right)$ → sublinear rate

↳ If f is μ-SC and also L-smooth, then

$\alpha_k = \dfrac{2}{1/\mu + 1/L}$ converges at a linear rate $O\left(\log(1/\varepsilon)\right)$

Dual decomposition:

$$\min_{x} \sum_{i=1}^{B} f_i(x_i)$$

s.t. $Ax = b$

$x = (x_1, x_2, \ldots, x_n)$

$\hookrightarrow x_i \in \mathbb{R}^{n_i}$

eg:

$f(x) = \frac{1}{2}(x_1^2 + x_2^2) + x_2 x_3$

$B_1 = x_1$

$B_2 = [x_2, x_3]$

$Ax = b$

$A = [A_1, \ldots A_B]$

← Broadcasting.

Scatter-Gather Algorithm:

Scatter ← $x_i^{(k)} = \arg\min_{x_i} f_i(x_i) + \nu_{k-1}^T A_i x_i$

Gather ← $\nu_i^{(k)} = \nu_i^{(k-1)} + \alpha_k \left(\sum_{i=1}^{B} A_i x_i^{(k)} - b \right)$