Augmented Lagrangian

Convergence of dual methods can be greatly improved by utilizing augmented Lagrangian. Start by transforming primal

$$\min_{x\in\mathbb{R}^n} f(x) + \frac{\rho}{2}\|Ax-b\|_2^2$$
$$\text{subject to } Ax = b$$

Clearly extra term $(\rho/2)\cdot\|Ax-b\|_2^2$ does not change problem

Assuming, e.g., A has full column rank, primal objective is strongly convex (parameter $\rho\cdot\sigma^2_{\min}(A)$), so dual objective is differentiable and we can use dual gradient ascent: repeat for $k=1,2,3,\ldots$

$$x^{(k)} = \underset{x\in\mathbb{R}^n}{\operatorname{argmin}}\ f(x) + (u^{(k-1)})^T Ax + \frac{\rho}{2}\|Ax-b\|_2^2$$
$$u^{(k)} = u^{(k-1)} + \rho(Ax^{(k-1)} - b)$$

Note step size choice $t_k = \rho$, for all k, in dual gradient ascent

Why? Since $x^{(k)}$ minimizes $f(x) + (u^{(k-1)})^T Ax + \frac{\rho}{2}\|Ax - b\|_2^2$ over $x \in \mathbb{R}^n$,

$$\begin{aligned} 0 &\in \partial f(x^{(k)}) + A^T\Big(u^{(k-1)} + \rho(Ax^{(k)} - b)\Big) \\ &= \partial f(x^{(k)}) + A^T u^{(k)} \end{aligned}$$

This is exactly the stationarity condition for the original primal problem; can show under mild conditions that $Ax^{(k)} - b$ approaches zero (primal iterates approach feasibility), hence in the limit KKT conditions are satisfied and $x^{(k)}, u^{(k)}$ approach optimality

Advantage: much better convergence properties

Disadvantage: not decomposable (separability compromised by augmented Lagrangian!)

ADMM

ADMM (Alternating Direction Method of Multipliers): go for the best of both worlds!

I.e., good convergence properties of augmented Lagrangians, along with decomposability

Consider minimization problem

$$\min_{x \in \mathbb{R}^n} \; f_1(x_1) + f_2(x_2) \;\; \text{subject to} \;\; A_1 x_1 + A_2 x_2 = b$$

As usual, we augment the objective

$$\min_{x \in \mathbb{R}^n} \; f_1(x_1) + f_2(x_2) + \frac{\rho}{2} \|A_1 x_1 + A_2 x_2 - b\|_2^2$$
$$\text{subject to } A_1 x_1 + A_2 x_2 = b$$

Write the augmented Lagrangian as

$$L_\rho(x_1, x_2, u) = f_1(x_1) + f_2(x_2) + u^T(A_1x_1 + A_2x_2 - b) + \frac{\rho}{2}\|A_1x_1 + A_2x_2 - b\|_2^2$$

ADMM repeats the steps, for $k = 1, 2, 3, \ldots$

$$\begin{aligned}
x_1^{(k)} &= \operatorname*{argmin}_{x_1 \in \mathbb{R}^{n_1}} L_\rho(x_1, x_2^{(k-1)}, u^{(k-1)}) \\
x_2^{(k)} &= \operatorname*{argmin}_{x_2 \in \mathbb{R}^{n_2}} L_\rho(x_1^{(k)}, x_2, u^{(k-1)}) \\
u^{(k)} &= u^{(k-1)} + \rho(A_1x_1^{(k)} + A_2x_2^{(k)} - b)
\end{aligned}$$

Note that the usual method of multipliers would have replaced the first two steps by

$$(x_1^{(k)}, x_2^{(k)}) = \operatorname*{argmin}_{(x_1, x_2) \in \mathbb{R}^n} L_\rho(x_1, x_2, u^{(k-1)})$$

Convergence guarantees

Under modest assumptions on f_1, f_2 (note: these do not require A_1, A_2 to be full rank), we get that ADMM iterates for any $\rho > 0$ satisfy:

- Residual convergence: $r^{(k)} = A_1 x_1^{(k)} - A_2 x_2^{(k)} - b \to 0$ as $k \to \infty$, i.e., primal iterates approach feasibility
- Objective convergence: $f_1(x_1^{(k)}) + f_2(x_2^{(k)}) \to f^\star$, where $f^\star$ is the optimal primal criterion value
- Dual convergence: $u^{(k)} \to u^\star$, where $u^\star$ is a dual solution

Note that we do not generically get primal convergence, but this can be shown under more assumptions

Practicalities and tricks

In practice, ADMM obtains a relatively accurate solution in a handful of iterations, but requires many, many iterations for a highly accurate solution. Hence it behaves more like a first-order method than a second-order method

Choice of ρ can greatly influence practical convergence of ADMM

- ρ too large $\rightarrow$ not enough emphasis on minimizing $f_1 + f_2$
- ρ too small $\rightarrow$ not enough emphasis on feasibility

Boyd et al. (2010) give a strategy for varying ρ that is useful in practice (but without convergence guarantees)

Like deriving duals, getting a problem into ADMM form often requires a bit of trickery (and different forms can lead to different algorithms)