We differentiate with respect to $x$ of the function we wish to minimize. We thus have all the components required of the optimal point of the reformulated problem:

$$
A^{(k)} = \left[ i \in \{1, \ldots, 2n\} | g_i^{(k)} \leq 0 \lor z_i^{(k)} > 0 \right].
$$

We now wish to consider only the elements of $z^{(k)}$ for those indices in $A^{(k)}$. Let this subvector be $z^{(k)_A}$. We also define the subvector $g^{(k)_A}$ similarly. Let $K^{(k)}$ be the submatrix formed by taking the indices of $K$ that are in the Cartesian product $A^{(k)} \times A^{(k)}$. Hence, $K^{(k)} > 0$.

We then wish to solve the following subproblem:

$$
-g^{(k)} = K^{(k)} \Delta z^{(k)}
$$

using conjugate gradient descent. For all $i = 1, \ldots, 2n$, we have:

$$
z_{(k+1)}^{(j)} = \begin{cases} 
(z_j^{(k)} + \alpha^{(k)} \Delta z^{(k)})_+ & \text{if } i = A_j^{(k)} \text{ for some } j \\
\frac{z_j^{(k)}}{\alpha^{(k)}} & \text{otherwise}.
\end{cases}
$$

Selecting $\alpha^{(k)}$ can be done with an Armijo-like step rule.

## Algorithm Pseudocode

Given $z^{(0)}$, $\lambda$, $\eta$.

```plaintext
k = 0.
repeat
1. $g^{(k)} = \left[ -A_b^{T} \lambda \eta + A^{T} A \lambda (1 - \eta) I + A_b + \lambda | \eta | \right] u / \lambda | \eta |.$
2. $A^{(k)} = \left[ i \in \{1, \ldots, 2n\} | g_i^{(k)} \leq 0 \lor z_i^{(k)} > 0 \right]$. 
3. Compute subvector $g^{(k)}$. 
4. Break if $\|g^{(k)}\| < \epsilon$. 
5. Compute submatrix $K^{(k)}$. 
6. Solve $K^{(k)} \Delta z^{(k)} = -g^{(k)}$ using conjugate gradient descent. 
7. Select a step size $\alpha^{(k)}$. 
8. for $i = 1, \ldots, 2n$:
   - $z_{(k+1)}^{(j)} = \begin{cases} \left( z_j^{(k)} + \alpha^{(k)} \Delta z^{(k)} \right)_+ & \text{if } i = A_j^{(k)} \text{ for some } j \\
   \frac{z_j^{(k)}}{\alpha^{(k)}} & \text{otherwise}.
\end{cases}$ 
9. $k = k + 1.$
```

## Future Work

- We want to adopt the heuristics glmnet uses to prematurely clamp variables.
- If $\eta = 1$, then $K$ might not be positive definite. So, we cannot use conjugate gradient descent to solve (5). We need to address this edge case.
- One of the other primary use cases of glmnet is in solving the logistic regression classification problem on very large data sets. We’d like to incorporate this functionality into our library.
- We would like to be able to leverage multiple machines to solve instances of the reformulated problem.

## Acknowledgements

I thank Hristov Paskov, my research partner, for his work in helping create the algorithm we use.